

AD-A263 482



2

GENDER AND ETHNICITY DIFFERENCES IN

MULTIPLE-CHOICE TESTING:

EFFECTS OF SELF-ASSESSMENT

AND RISK-TAKING PROPENSITY

BY

EILEEN PATRICIA WILLIAMS, M.A.

A Thesis Submitted to the Graduate School
in partial fulfillment of the requirements

for the Degree

Master of Arts

Major Subject: Psychology

SDTIC
ELECTE
APR 30 1993
S E D

New Mexico State University

Las Cruces, New Mexico

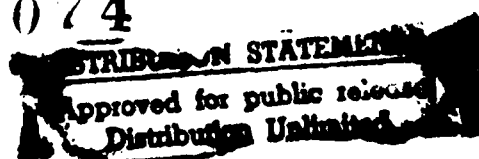
May 1993

93 4 21 074

93-08626



144P8



REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

MAY 1993

3. REPORT TYPE AND DATES COVERED

FINAL - THESIS

4. TITLE AND SUBTITLE

Gender and Ethnicity Differences in Multiple-Choice Testing: Effects of Self-Assessment and Risk-Taking Propensity

5. FUNDING NUMBERS

6. AUTHOR(S)

CPT. Eileen P. Williams

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Attached are two copies of thesis from New Mexico State University. Thesis submitted in compliance w/ USAF Student Handbook

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

See Thesis

14. SUBJECT TERMS

15. NUMBER OF PAGES

129

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

"Gender and Ethnicity Differences in Multiple-Choice Testing: Effects of Self-Assessment and Risk-Taking Propensity," a thesis prepared by Eileen Patricia Williams in partial fulfillment of the requirements for the degree, Master of Arts, has been approved and accepted by the following:

L L Ames

Lynford L. Ames
Dean of the Graduate School

Darwin P. Hunt

Darwin P. Hunt
Chair of the Examining Committee

4-14-93

Date

Committee in charge:

Dr. Darwin P. Hunt, Chair

Dr. Dennis B. Beringer

Dr. Melchor Ortiz, Jr.

Dr. Laura A. Thompson

Statement A per telecon
Capt Jim Creighton TAPC-OPB-D
Alexandria, VA 22332-0411

NWW 4/28/93

Accession For	
NTIS	CRA&I <input checked="" type="checkbox"/>
DTIC	TAB <input type="checkbox"/>
Unannounced <input type="checkbox"/>	
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

VITA

29 April 1957	Born in Aurora, Colorado
4 June 1975	Enlisted U.S. Army
16 April 1982-Present	Commissioned Officer U.S. Army
1985	B.S. Sociology, University of Maryland, European Division, Wuerzburg, Germany
1991	M.A. Counseling, Webster University, Ft. Bliss, Texas

PROFESSIONAL AND HONORARY SOCIETIES

Association of the United States Army

National Stenomask Verbatim Reporters Association

American Association for Counseling and Development

Human Factors Society

ACKNOWLEDGEMENTS

I'd like to thank the members of my committee for their valuable comments and suggestions. Special thanks go to Dr. Darwin P. Hunt; his guidance, and untiring efforts in searching out "the truth" are to be commended.

I am grateful to Dr. Hunt, and Holt, Rinehart, and Winston, for allowing me to use copyrighted material; specifically the Multiple-Choice Self-Assessment Answer Sheet (1983, 1990), and the Opinion II Questionnaire (1964).

I am especially indebted to Dr. Melchor Ortiz for his advice, help, patience, and kindness during my "thesis experience." His entire ESTAT crew (Gloria, Bernice, and Mary) were more than supportive; they were friends.

Thanks also go to Dru Sharp who saved me from having to type this thesis myself.

I am grateful to my husband, Jim, and to my two sons, Stanley and Jimmy who have given me love and moral support for yet another personal and professional endeavor; they've spent more time without me, and have eaten more fast food than should be allowed by law.

And last, but in no way least, I would like to thank Jesus Christ. Because of his love and strength, and

presence in my life, I have been able to accomplish many difficult and sometimes insurmountable tasks (i.e., Quantitative Methods I and II). It is true, I can do all things through Christ which strengtheneth me.

ABSTRACT

GENDER AND ETHNICITY DIFFERENCES IN
MULTIPLE-CHOICE TESTING:
EFFECTS OF SELF-ASSESSMENT AND RISK-TAKING PROPENSITY

BY

EILEEN PATRICIA WILLIAMS, M.A.

Master of Arts in Psychology
New Mexico State University
Las Cruces, New Mexico, 1993
Dr. Darwin P. Hunt, Chair

The following thesis attempted to (a) test the robustness of Hassmen and Hunts' (1990) findings regarding the self-assessment technique; this time considering Hispanic test performance, and (b) determine if the self-assessment process was related to subjects' risk-taking propensity.

Two-hundred and forty college students enrolled in Psychology 201 classes at New Mexico State University were given a fifty item multiple-choice test. Subjects marked their answers on a usability assessment answer

sheet, a self-assessment answer sheet, or a standard multiple-choice answer sheet.

The usability and self-assessment answer sheets are modified forms of the standard multiple-choice answer sheet. The usability assessment answer sheet has a section where the respondent assesses the usefulness of the information contained in each test item. The self-assessment answer sheet has a section where the respondent assesses the level of sureness of each answer. Both types of assessment are done immediately following selection of an answer.

Each subject was also given a risk propensity test following the multiple-choice test.

The results failed to support the hypothesis that engaging in self-assessment after each question would enhance females' and Hispanics' test performance. Additionally, females who self-assessed did not have less conservative risk propensity scores than females who did not self-assess.

An analysis of the data revealed that Non-Hispanic males' and females', and Hispanic males' multiple-choice test scores did not differ significantly. However, Hispanic females' test scores were statistically lower than these three groups.

There was no significant difference among the means of the treatments for Non-Hispanics or Hispanics. However, there were differences in the treatments between the ethnicities.

Non-Hispanics who were tested with the usability assessment treatment scored significantly higher than Hispanics from all three treatment groups. Non-Hispanics who self-assessed, and those tested without self-assessment scored significantly higher than Hispanics who made usability assessments and Hispanics who self-assessed.

There was no significant difference among the scores of Non-Hispanics who self-assessed, Non-Hispanics who did not self-assess, and Hispanics who did not self-assess.

While the risk scores for Hispanic females (\bar{M} = 78.1) and Non-Hispanic males (\bar{M} = 62.9) tested without self-assessing were significantly different from each other, neither one alone was different from the rest of the groups.

TABLE OF CONTENTS

	PAGE
List of Tables	xi
List of Appendix Tables	xii
List of Figures	xiii
List of Appendix Figures	xiv
Chapter	
1. INTRODUCTION	1
Multiple-Choice Tests	1
Multiple-Choice Tests and Gender Bias	6
Multiple-Choice Tests and Hispanics	11
Self-Assessment	17
Risk-Taking	23
Pilot Study	27
2. HYPOTHESIS	31
Method	31
Subjects	31
Design/Instruments	32
Procedure	34
Results	36
3. DISCUSSION	42
REFERENCES	49
APPENDICES	54
A. PILOT STUDY	54
B. SAMPLE SAT MULTIPLE-CHOICE TEST (50 ITEMS) AND ANSWER KEY	70

C.	STANDARD MULTIPLE-CHOICE ANSWER SHEET . . .	82
D.	MULTIPLE-CHOICE SELF-ASSESSMENT ANSWER SHEET (1983 VERSION)	85
E.	RISK-TAKING TEST	88
F.	VERBAL INSTRUCTIONS GIVEN FOR SAMPLE SAT TEST	98
G.	VERBAL INSTRUCTIONS GIVEN FOR RISK-TAKING TEST	100
H.	SELF-ASSESSMENT INSTRUCTIONS	102
I.	SELF-ASSESSMENT ANSWER SHEET (1990 VFRSION)	104
J.	USABILITY ANSWER SHEET	107
K.	INSTRUCTIONS FOR USABILITY ANSWER SHEET . .	110
L.	INSTRUCTIONS FOR MULTIPLE-CHOICE ANSWER SHEET	112
M.	SUPPLEMENTARY FIGURES AND TABLES	114

LIST OF TABLES

TABLE	PAGE
1 1988-89 LAUSD and National SAT Scores: A Comparison Between Ethnicities	13
2 Scoring Matrix for the Self-Assessment Answer Sheet	19
3 Sample Sizes for Each Ethnicity, Gender and Treatment	32
4 Means and Variances for Number of Correct Responses for Treatment, Ethnicity, and Gender Based on 20 Observations Per Group . . .	37
5 Ethnicity and Gender Mean Pairings From Protected Least Significant Difference Comparisons. Means with the Same Letter are not Significantly Different	39
6 Ethnicity and Treatment Mean Pairings From Protected Least Significant Difference Comparisons. Means with the Same Letter are not Significantly Different	39
7 Means and Variances for Risk Scores for Treatment, Ethnicity, and Gender Based on 20 Observations Per Group	41

LIST OF APPENDIX TABLES

APPENDIX TABLE	PAGE
A1 Sample Sizes for Each Ethnicity, Gender, and Treatment No Self-Assessment-NOSA, Self-Assessment-SA . . .	56
A2 Means for Number of Correct Responses, and Sample Sizes for Ethnicity, Gender, and Treatment No Self-Assessment-NOSA, Self-Assessment-SA . . .	62
A3 Means for Risk Scores and Sample Sizes for Each Ethnicity, Gender, and Treatment No Self-Assessment-NOSA, Self-Assessment-SA . . .	63
M1 Means and Variances of Number of Correct Responses for Treatment, Ethnicity, and Gender Based on 20 Observations Per Cell	115
M2 ANCOVA Table Showing <u>p</u> Values Calculated for Number of Correct Responses For Gender, Ethnicity, and Treatment Based on 20 Observations Per Cell	116
M3 Mean Risk Scores For Treatment, Ethnicity, and Gender Based on 20 Observations Per Cell . . .	117

LIST OF FIGURES

FIGURE	Page
1. Mean Number Correct For Non-Hispanics and Hispanics By Gender	38
2. Mean Number Correct For Non-Hispanics and Hispanics By Treatment	38

LIST OF APPENDIX FIGURES

APPENDIX FIGURE	PAGE
A1 Mean Number Correct for Each Treatment for Native Americans, Hispanics, and Non-Hispanics by Gender	64
M1 Mean Number Correct for Each Treatment for Non-Hispanic and Hispanic Females	118
M2 Mean Number Correct for Each Treatment for Non-Hispanic and Hispanic Males	118
M3 Median Number Correct for Each Treatment for Non-Hispanic and Hispanic Females	119
M4 Median Number Correct for Each Treatment for Non-Hispanic and Hispanic Males	119
M5 Variances for Mean Number Correct for Each Treatment for Non-Hispanic and Hispanic Females	120
M6 Variances for Mean Number Correct for Each Treatment for Non-Hispanic and Hispanic Males	120
M7 Standard Deviations for Mean Number Correct for Each Treatment for Non-Hispanic and Hispanic Females	121
M8 Standard Deviations for Mean Number Correct for Each Treatment for Non-Hispanic and Hispanic Males	121
M9 Mean Risk Score for Each Treatment for Non-Hispanic and Hispanic Females	122
M10 Mean Risk Score for Each Treatment for Non-Hispanic and Hispanic Males	122
M11 Median Risk Score for Each Treatment for Non-Hispanic and Hispanic Females	123
M12 Median Risk Score for Each Treatment for Non-Hispanic and Hispanic Males	123

M13	Variances for Mean Risk Score for Each Treatment for Non-Hispanic and Hispanic Females	124
M14	Variances for Mean Risk Score for Each Treatment for Non-Hispanic and Hispanic Males	124
M15	Standard Deviations for Mean Risk Score for Each Treatment for Non-Hispanic and Hispanic Females	125
M16	Standard Deviations for Mean Risk Score for Each Treatment for Non-Hispanic and Hispanic Males	125
M17	Mean GPA for Each Treatment for Non- Hispanic and Hispanic Females	126
M18	Mean GPA for Each Treatment for Non- Hispanic and Hispanic Males	126
M19	Median GPA for Each Treatment for Non- Hispanic and Hispanic Females	127
M20	Median GPA for Each Treatment for Non- Hispanic and Hispanic Males	127
M21	Mean Age for Each Treatment for Non- Hispanic and Hispanic Females	128
M22	Mean Age for Each Treatment for Non- Hispanic and Hispanic Males	128
M23	Median Age for Each Treatment for Non- Hispanic and Hispanic Females	129
M24	Median Age for Each Treatment for Non- Hispanic and Hispanic Males	129

Chapter 1

INTRODUCTION

This study was designed to investigate gender and ethnicity differences in multiple-choice testing using Hunt's (1984) self-assessment (SA) technique. Risk-taking propensities were also examined in male and female subjects to determine if a relationship exists between self-assessment and risk-taking.

Multiple-Choice Tests

According to Echternacht (1972), the method used most widely for measuring scholastic ability and achievement in our educational system is the multiple-choice examination. Multiple-choice tests are used more frequently than any other test because more items can be administered in a given period of time using this method than by any other method requiring a more complicated response, and the cost for scoring the test is less. Aiken (1987) claims that multiple-choice tests have the advantages of:

1. Versatility. They measure both simple and complex objectives at almost all grade levels and in all subject areas;

2. Sampling more adequately. They can sample the domain of abilities more satisfactorily than essay items and almost all other objective items;
3. Being less susceptible than true-false items to both guessing and response sets, and greater reliability than true-false items;
4. Objectivity in scoring. They can be scored accurately and rapidly by almost anyone; and
5. Objectivity and ease in item analysis.

Unfortunately, there are also disadvantages associated with this test format. Hassmen and Hunt (1990) discuss them in detail in their study on reducing gender bias in multiple-choice testing using self-assessment. They are:

1. Difficulty associated with constructing good items, e.g., items which measure higher-order objectives that have an adequate number of parallel alternatives. This process is also very time consuming.
2. Response times are greater for multiple-choice items as compared with true-false items.
3. They may sample the domain of knowledge less completely than essay questions; and

4. They emphasize recognition of the correct answer rather than recall.

Critics such as Hoffman (1962) believe that multiple-choice items are concerned only with the answer and not with the quality of thought behind it or the skill with which it is expressed. Hoffman also asserts that the multiple-choice format allows rapid readers an unfair advantage over creative, more profound individuals.

Even though the criticisms of multiple-choice tests are valid, there appears to be no other viable alternative because class sizes have increased over the years, it is more costly to develop and grade other types of tests, and the subjectivity involved in grading other types of tests would probably outweigh their benefits.

Aiken (1987) predicts that the use of multiple-choice tests will increase in the future and that we may have to learn to live with their shortcomings.

A way to improve information gained from multiple-choice tests and to overcome negative features may be to improve scoring methods (Hassmen & Hunt, 1990). Many multiple-choice tests are scored by simply counting the number of correct responses. This method does not account for guessing. Other tests such as the Scholastic Aptitude Test (SAT) utilize a formula whereby guessing is

penalized. The College Board decided that to encourage guessing was educationally unsound and morally improper (Angoff, 1971). However, even formula scoring has been criticized as yielding over-corrected scores when test takers are less familiar with the test material and under-corrected scores when they are more familiar with it (Hassmen & Hunt, 1990). Glass and Wiley (1964) reported that the correction formula decreases reliability while Lord (1963) has shown that it increases validity.

Slakter (1968) investigated scoring methods which penalized test takers for guessing. Test directions were administered which warned students against guessing, and scoring formulas included a "penalty for guessing." Slakter found that "do not guess" instructions caused certain test takers to take fewer risks and tended to waste partial information. High risk-takers did not appear to be affected. Slakter modified the "do not guess" instructions to encourage low risk-takers to utilize their partial information, but he found that some students were unable to discern between complete, partial, and no information and these students were penalized more than others.

Slakter's (1968) findings suggest that examinees should not be discouraged from guessing when taking

multiple-choice tests. Wood (1976) asserts that guessing contributes to the validity of the measurement.

Shuford, Albert, and Massengill (1966) propose confidence-weighting as an alternative to conventional scoring methods. Test takers assign probability weights to each alternative on each item. The weights are determined by subjects' certainty that the option is the correct one (Rippey & Voytovich, 1985). Anderson (1982) reports that confidence testing which requires examinees both to make a correct response and to express a level of confidence in the correctness of the response provides some advantages. They are:

1. Increased reliability of the test;
2. Examinees pay more attention to the multiple-choice alternatives;
3. More diagnostic information becomes available; and,
4. Pre-and post examination tension is reduced, leading to happier examinees.

Bokhorst (1986) administered a multiple-choice test using the confidence approach. Results showed that confidence weighting did not improve the validity of the test and was slightly inferior to the conventional scoring method. These findings are similar to those reported by Hopkins et al. (1973).

Echternacht (1972) proposes that when using confidence weighting too little is gained at too great a cost, while Shuford et al., (1966) state that the method has both theoretical and practical advantages in that it assesses the realism of self-perceived knowledge. Swineford (1938) identified a personality variable that differed between males and females in confidence weighting. Males tended to gamble significantly more often than did females on test responses; and both males and females tended to gamble more on unfamiliar material than familiar material. Jacobs (1971) questioned the use of confidence weighting based on results that showed scoring procedure tends to be contaminated by individual differences in personality.

Arguments for and against different types of scoring methods continue.

Multiple-Choice Tests and Gender Bias

Another major criticism of multiple-choice testing is its alleged built-in gender bias, favoring males over females (Bolger & Kellaghan, 1990; Hassmen & Hunt, 1990).

Rosser (1989) asserts that bias can be expressed in four ways:

1. In test content; males are depicted more often than females and females are shown in lower status or stereotyped roles.

2. In test context; questions are set in experiences more familiar to one sex than the other. Females tend to prefer questions with aesthetic-philosophical and human relations content while males prefer questions dealing with science or practical affairs.
3. In test validity; females' academic abilities are under-predicted by test scores while males' are over-predicted; and,
4. In test use; females' access to educational opportunities are diminished by an institution's reliance on a test that under-predicts their ability.

Different theories exist to account for this gender difference in multiple-choice testing (Hassmen & Hunt, 1990). They include:

1. "Test-wiseness." Hassmen and Hunt (1990) define test-wiseness as "the ability to respond advantageously to multiple-choice items containing extraneous clues and, therefore, to obtain credit without knowledge of the subject matter being tested" (p. 6).
2. Cognitive differences in the way males and females deal with multiple-choice questions, and

3. Greater omission rates for females compared with males.

Maccoby and Jacklin (1974) conducted extensive research on intellectual performance differences between males and females. They found that males outperform females in mathematical and spatial subjects, and that females have greater verbal abilities. Maccoby and Jacklin (1974) also suggest that females are lower in self-confidence than males in achievement settings such as testing.

Campbell and Fiske (1959) assert that variance in test scores may be due to the form of the test used and individual characteristics that the test is designed to measure. Bolger and Kellaghan (1990) expect student characteristics such as cognitive style, test-wiseness, and risk-taking to interact with measurement method. In their 1990 study they found males performed significantly better than females on multiple-choice tests compared to free response or essay tests. These differences were evident in two types of mathematics exams. Females performed relatively better on the essay type examination. Bolger and Kellaghan (1990) attributed females' poorer performance on the multiple-choice test to their inability to deal with novel situations and a lower propensity to guess.

Skinner (1983) discovered that females changed their answers on multiple-choice tests twice as often as males. He suggests that this behavior may have a negative effect on the performance of timed tests. Pascale (1974) found that even though males did not change their answers as often as females, when they did they were more successful.

Females were also found to have higher omission rates on multiple-choice tests than males, especially with mathematical questions (Ben-Shakhar & Sinai, 1991). Ben-Shakhar and Sinai discovered that females failed to answer more questions than males even on subtests which showed no significant differences in performance between genders, and when given permissive instructions that encouraged guessing. Rosser (1989) asserts that this tendency on the part of females to omit more than males may indicate that females have more difficulty with multiple-choice type tests than males.

Hassmen and Hunt (1990) acknowledge gender differences exist in multiple-choice testing (page 20).

Findings alleging gender bias in multiple-choice testing have serious ramifications for our educational system and society as a whole. Not only are multiple-choice test scores being used to predict such things as academic success, they are considered for determining

which students are accepted into college programs and for awarding scholarships as well.

One of the most widely used and controversial multiple-choice tests is the Scholastic Aptitude Test (SAT). The test consists of six parts which test students' verbal and mathematical reasoning abilities. The student is given 30 minutes to complete each section; the entire test takes three hours. The SAT was administered for the first time in 1926 by the College Board in order to standardize college entrance examinations. Since then, over two million students each year take the SAT to satisfy college entrance requirements (Angoff, 1971). Scores are used by colleges to measure a student's aptitude for college work, to predict the student GPA during their freshman year, and to assist the student in selecting an academically appropriate college based on their score (Cruise & Trusheim, 1988). Many critics feel the SAT is overrated and doesn't assist colleges or students in any of these claims.

Prior to 1975, females earned higher scores than males on the verbal portions of the SAT. Females' math scores were much lower than males' math scores. Since 1975, males have scored higher on the verbal portions of

the SAT and continued to outscore females on the math portions (Angoff, 1971).

Clark and Grandy (1984) compared SAT test performance in 1972, with 1983, and found declines in the average SAT verbal scores from 454 to 430 (24 points) for males and from 452 to 420 (32 points) for females; the decline in average SAT mathematical scores since 1972 also were greater for females, from 461 to 445 (16 points) than for males 505 to 493 (12 points).

According to Hassmen and Hunt (1990), the mean SAT score overall for females is 60 points lower than for males. This difference in scores could mean that fewer females will receive scholarships to prestigious universities.

Multiple-Choice Tests and Hispanics

Test performance differences have also been studied extensively with respect to other minorities; mainly Blacks (Goldman & Newlin-Hewitt, 1975). According to Temp (1971), these investigations have proven to be valuable, but have not addressed the issue as it concerns other minority subgroups. Further, Temp (1971, p.247) states, "Most investigations have dealt solely with black students and then the generalizations have been

extrapolated to other minorities (i.e., Mexican Americans, the disadvantaged, low income females, etc.)."

These generalizations, especially if applied to Hispanics, can be considered invalid because major issues such as socioeconomic, cultural, and linguistic factors are not taken into account (Goldman & Newlin-Hewitt, 1975).

Studies regarding test performance differences have shown that even though Hispanics have increased their SAT scores in the past decade, an "ethnic gap" still exists between them and Non-Hispanics (Isonio, 1990).

For the purposes of this study, the term Non-Hispanics is used to refer to those persons that are considered as White and not Hispanic (M. Loustau, personal communication, 5 March 1993).

The Los Angeles Unified School District (LAUSD) administered the SAT to 10,775 high school students during the 1988-89 school year and compared their scores to the national average (Isonio, 1990); (see Table 1). Differences between Hispanics' scores and Non-Hispanics' scores are clearly apparent.

As mentioned above, there are a number of factors which could be responsible for the academic underachievement of Hispanics as compared to Non-

Hispanics. According to Mestre (1988), Hispanic culture has an effect on cognitive performance. Most studies have focused on familism and how it may affect cognitive performance.

Table 1

1988-89 LAUSD and National SAT Scores: A Comparison Between Ethnicities

ETHNICITY	LAUSD		NATIONAL	
	Verbal/Math		Verbal/Math	
Non-Hispanic	455	504	446	491
Hispanic	378	428	380	427

Familism can be defined as the relative importance of family members in determining an individual's values, goals, and orientation (Mestre, 1988).

Grebler, Moore, and Guzman (1970) have argued that the Hispanic family obstructs intellectual development because family needs are placed above individual needs.

Schwartz (1971) found that Hispanics who are more independent of their families attain greater educational achievements than Hispanics who retain closer family ties.

Aiken (1979) asserts that while Hispanic and Non-Hispanic parents may not differ in the value they place on education for their children, Hispanic parents tend to

encourage their male children to pursue advanced education more than their female children.

Mestre (1988) contends that there is a clear difference between Non-Hispanic and Hispanic family values in one area: Hispanic parents are more traditional in their attitudes toward gender roles than Non-Hispanic parents are; Hispanic girls are encouraged to put their future families ahead of their career and educational pursuits.

Although research evidence shows that Hispanic children are more likely to do their homework than Non-Hispanic children, and that Hispanic parents are very supportive of their children's education, MacCorquodale (1988) argues that Hispanic parents have difficulty in translating their encouragement and support into concrete actions. This may be due to their limited educational background. Evidence also exists which shows that culture directly affects cognitive performance; specifically reading comprehension. A lack of language proficiency can also affect cognitive performance.

Duran (1983) proposes that differences in test scores of Hispanics and Non-Hispanics are a result of true differences in skill development as well as cultural and language differences. He contends that tests such as the SAT lack in providing diagnostic information on

students' learning aptitudes that can be used to prescribe specific learning interventions (Duran, 1988). Results of his experiment are consistent with those of Goldman and Duran (1988), which showed that bilinguals have greater difficulty in maintaining an accurate working memory for information presented in their less familiar language.

Imposing a time limit during testing may have an effect on test performance for Hispanics. Younkin (1986) studied the effects of increased testing time on the performance of 659 native and non-native Hispanic speakers of English. Native speakers showed no improvement with increased time, but non-native speakers improved up to 1/3 standard deviation with increased time (Younkin, 1986).

Schmitt and Dorans (1987) also examined the effects of timing during testing. They analyzed the results of a 1983 SAT test; specifically the ten analogy items located at the end of the forty-five-verbal-item section of the SAT. They compared Hispanics and Non-Hispanics of equal ability and found that all ten analogy questions were reached by a higher proportion of Non-Hispanic examinees than Hispanic examinees (Schmitt & Dorans, 1987).

Llabre and Froman (1987; 1988) also conducted studies which compared Hispanic and Non-Hispanic college

students with respect to time allocation to cognitive test items. Both of their studies indicated that Hispanics take longer than Non-Hispanics of equal ability in responding to both verbal and nonverbal test items; if time is not restricted, the two groups do not differ significantly in test performance (Llabre & Froman).

Finally, Schmitt (1988) conducted a differential item functioning (DIF) study which identified factors that differentially affect the performance of Hispanics on items and result in underestimating their potential and competence. Schmitt studied the effects true and false cognates would have on Hispanic test performance. True cognates are words with a common root in both English and Spanish, and false cognates appear to have the same root in English and Spanish but in reality have quite different meanings in each language (Schmitt). Schmitt found that true cognates tended to favor Hispanic examinee item functioning and false cognates impeded their performance.

Schmitt (1988) also studied the effects of homographs on Hispanic examinee item functioning. A homograph is a word with the same spelling as another word but having different meanings and word roots. Results showed that homographs impeded the performance of Hispanic examinees.

Hispanics have been shown to score lower than the majority population on tests which assess academic aptitude and achievement. As with females, low scores on such tests as the SAT could result in Hispanics receiving fewer scholarships which would enable them to advance their education.

Self-Assessment

Hunt's (1982, 1984) self-assessment technique offers an alternative which may reduce gender and ethnicity differences in multiple-choice testing.

According to Hunt, the standard multiple-choice test encourages the test taker to guess even though the test taker may have no feeling of confidence in his answer. Hunt's method allows the test taker to indicate doubt or sureness about each answer and is more similar to the way in which individuals use knowledge to make decisions in day-to-day life situations (Hunt, 1991). If a test taker assesses himself too low then he may fail to reach his full potential. Conversely, if he assesses himself too high he suffers the consequences of too many errors, and he lacks the knowledge he thought he possessed (Hunt, 1991).

Self-assessment possesses two unique advantages. First, it provides a measurement of a test taker's

"usable" knowledge. Hassmen and Hunt (1990, p. 8) define usable knowledge as "that knowledge about which a person is sufficiently sure so that the knowledge will be used in making decisions, solving problems, and in selecting and executing actions." This concept has important implications for learning and testing. Similar self-assessment testing methods were evaluated in the Los Angeles school system with overwhelming favorable results. Students profess that it is more fair than standard multiple choice testing, and reduces test anxiety. Teachers indicate that it gives better information to help students learn and is seen as "a more accurate measure of the knowledge base of the individual student" (Hunt, 1991, p. 2).

The second advantage of self-assessment testing is that it can "detect and identify topics about which students are misinformed" (Hunt, 1991, p. 2). If a test taker is sure of the correctness of his answer, but is wrong then he may be considered misinformed. The self-assessment technique can also indicate if a test taker is fully informed, partially informed or uninformed.

Hunt has conducted extensive research using the self-assessment technique and has reported significant findings in learning and in training (Hunt, 1982, 1984; Sams, 1989).

Hunt (1982; 1984) modified the standard multiple-choice answer sheet by adding a section after each question which enables the test taker to express their level of sureness in their answer. There are five choices. They range from "Almost a Guess," through "Neutral," to "Almost Certain." Points are lost or gained depending upon the correctness of the answer and the accuracy of the self-assessment (Hassmen & Hunt, 1990). Credit is given for correct answers, with more credit given if the test taker is "Sure" of its correctness. Some credit is even given for incorrect answers if it is indicated that the test taker was not sure at all. However, a penalty is given for answers that are incorrect and which the test taker marked "Sure" (see Table 2).

Table 2

Scoring Matrix for the Self-Assessment Answer Sheet

Answer	Almost a Guess	Probable Guess	Neutral	Fairly Certain	Almost Certain
Correct	+10	+27	+37	+45	+50
Wrong	+5	-4	-16	-32	-60

This scoring method yields a percentage self-assessment score which can be described as an overall index of the

accuracy with which each student assessed the correctness of their answers (Hunt, 1991).

Hassmen and Hunt (1990) provide three reasons why self-assessment should be applied to multiple-choice testing. They are:

1. To make the multiple-choice test more accurate and comprehensive in measuring the knowledge of the test taker,
2. To give extra credit to the person who not only knows the topic being tested, but is sure of that knowledge, and
3. To allow test takers to express their doubt or certainty about the answers they select which may have some beneficial effects regarding issues of gender bias, cultural bias, test anxiety, etc.

Hassmen and Hunt (1990) conducted research to determine whether making self-assessments regarding the correctness of answers affected a test takers' score, and whether there were, in fact, differences between the scores of males and females using, or not using the self-assessment technique. The SAT test was used for reasons previously discussed. They selected 50 "gender equal" items (questions referred to males and females in an equal way) and included 10 mathematical and 40 verbal

items. Each item had five alternatives with only one alternative being correct. In their study, one male and one female group ($n=30$ each) answered questions using the standard multiple-choice answer sheet and one male and one female group ($n=30$ each) answered the same questions using the self-assessment answer sheet.

Hassmen and Hunt (1990) found a significant difference in the number of correct answers for females who self-assessed compared to females who did not self-assess. Females who self-assessed showed higher scores (mean number correct) compared to females who did not (27.7 vs. 23.9). There were no significant differences between males' scores (29.70 vs. 29.2). The "gap" between males' scores and females' scores was lessened when self-assessment was used.

Findings did not prove that males were more accurate in their self-assessments than females (74.0% versus 73.1%), but males did score a higher sure-and-correct score (mean number correct) (30.7) than did females (22.9). Hassmen and Hunt (1990) speculated that either males are better able to identify a correct response once it has been selected, or possibly female test takers feel more stress than males when taking tests.

Sams (1986), who used only female subjects, found that the performance of subjects was positively affected

simply by asking them to assess the correctness of their answers.

Palmer (1990) also studied gender differences in multiple-choice testing. Subjects were given a test similar to the SAT; half of the subjects answered the questions on the conventional multiple-choice answer sheet, and the remaining subjects answered questions using the self-assessment answer sheet. Palmer was interested in the effect anxiety had on cognitive performance and whether self-assessing would reduce anxiety. Palmer generated anxiety by reading different test instructions to three different groups. The instructions were intended to cause low, medium, or high levels of anxiety. Subjects were required to stop at question 34 on the test and assess their levels of anxiety by answering the Affect Adjective Checklist. He found significant gender differences in perceived anxiety; females reported higher levels of anxiety across all conditions than males. Results failed to support the hypothesis that engaging in self-assessment would enhance performance by reducing anxiety.

It should be noted, however, that Palmer's study was not an exact replication of the Hassmen and Hunt (1990) study. For example, Palmer used 60 SAT questions; 30 mathematical and 30 verbal whereas Hassmen and Hunt

used only 10 mathematical and 40 verbal. As discussed earlier, it has been shown that females score much lower on mathematical questions than males (Angoff, 1971). Palmer's test contained a higher proportion of mathematical questions than did Hassmen and Hunts' test and this may have produced the difference in findings.

Unfortunately, Hassmen and Hunt, Sams, and Palmer did not collect data concerning ethnicity and self-assessment.

Risk-Taking

In his research, Palmer (1990) hypothesizes that performance differences between males and females on the SAT are the result of gender differences in response to conditions that elicit anxiety. According to evolutionary theory, risk reduction is of paramount importance to females since they are responsible for giving birth to and caring for their offspring. High risk behaviors would be hazardous to fitness.

Palmer (1990) suggests that the structure of the multiple-choice test imposes a perceived risk on the subject. For example, the subject must select a response and claim, without explanation, that it is correct. This causes some degree of anxiety. In his study, Palmer found that female subjects reported higher levels of

anxiety than males when tested with the multiple-choice format. This may be because of the risk associated with choosing an answer that may or may not be correct.

Maccoby and Jacklin (1974) found that, in child rearing, boys are reinforced for and girls are discouraged from engaging in risk-taking behaviors.

Risk-taking propensity of females should be of interest to educators, especially if it has a negative effect on females' performance on examinations such as the SAT.

What does the literature have to say about females and risk-taking? According to Rosser (1989) females are less likely to be risk-takers and less likely to guess at the right answer; they attribute this largely to their upbringing, socialization and earlier education. They found in a study using a science assessment test, the National Assessment of Educational Progress, that girls more than boys used the "I don't know response" especially for perceived masculine items. Rosser (1989) suggests that their unwillingness to take risks may lead females to avoid giving a definite answer.

Plax and Rosenfeld (1976) discovered a correlation between certain personality variables and subjects' responses to risk tests. They found these variables correlated significantly with risky decision making.

They assert that as an individual's decision making became more risky, he or she exhibited behaviors associated with masculinity.

Kogan and Wallach (1964) studied sex differences in risk-taking and found that females had less confidence in their probability estimates and possessed narrower category widths. Category width can be explained as a type of cognitive risk measure. According to Kogan and Wallach (1964), a person's possession of broader or narrower category boundaries evidently involves a preference for errors of inclusion or exclusion. They found that some subjects would risk including instances not belonging to a category, rather than risk leaving them out while other subjects preferred to leave a few "correct" instances outside the category, rather than risk including any instances that might not belong (Kogan & Wallach, 1964). A narrower category width suggests conservatism. Kogan and Wallach (1964) propose that "feminine conservatism is learned through fear of punishment in subjectively ambiguous situations, but that when a situation is perceived as highly certain, a counterphobic release of boldness seems to occur" (p.12).

Slovic (1964) suggests category width may be a valid tool to use in evaluating risk propensity. Results of testing in Kogan and Wallach's studies found females

didn't display as high a degree of certainty as often as men, but when they were certain they would take high risks.

Hudgens and Fatkin (1985) tested sex differences in risk-taking behavior using a computer-generated and controlled task. They used military men and women as their subjects. The task required the subjects to decide whether to send his or her tank across a minefield when the only information available was the number of visible mines. They confirmed their hypothesis that males were greater risk-takers than females. They also found that the females took longer to make decisions.

Finally, Ben-Shakhar and Sinai (1991) found that males took greater risks while being tested using the multiple-choice format than females. That is, they guessed more often even though they knew they could be penalized for such behavior.

As can be concluded from the preceding review, gender and ethnicity differences exist in multiple-choice testing. There are also gender differences in risk-taking propensity. However, an extensive review of the literature on risk-taking revealed no information regarding risk-taking differences between ethnicities.

Hunt's self-assessment technique may facilitate risk-taking for females when taking multiple-choice tests

by providing a situation in which females may express the levels of their certainty or uncertainty. These females may then be able to adopt a higher risk-taking propensity than females who are tested with the usual multiple-choice format. Results should show higher test scores for females who self-assess than for females who do not.

As previously mentioned, making self-assessments regarding the correctness of answers may also have some beneficial effect regarding the issue of cultural bias. If so, the "gap" between Hispanics' scores and Non-Hispanics' scores should be lessened.

Pilot Study

A pilot study was conducted to select suitable methods, procedures, and testing materials so that an improved study could be performed to determine whether: (a) using self-assessment during testing improves a test taker's score i.e., the number correct; (b) females who self-assess achieve a higher number correct than females who don't self-assess; and (c) the risk scores for females who self-assess are less conservative than the risk scores of females who do not self-assess (see Appendix A).

The overall design of the experiment may be described as a between-subjects, 2 X 2 factorial, with

the independent variables being : Self-Assessment - SA (with) and NOSA (without), and Gender - Male (M) and Female (F). Information concerning age, GPA (high school or college freshman), and ethnicity (White and Black Non-Hispanic, Hispanic, Native American) was obtained from each subject.

The dependent variable was test performance measured in number correct. The risk propensity score was used as a tool to try to interpret the hypothesized difference in scores. The alpha level was set at 0.10 for the purposes of the pilot study only.

An Analysis of Covariance (ANCOVA) revealed a three-way interaction among gender, self-assessment, and ethnicity with a probability of error equal to 0.07. Although this value is not significant when compared to the more commonly used .05 level, it suggests that something of interest might be occurring. GPA and age were used as the covariates. Effects of self-assessment were different depending on gender and ethnicity. Analyzing the data further using the protected Least Significant Difference procedure revealed that self-assessment appears to have had a positive impact for Hispanic females and Hispanic and Native American males. Non-Hispanics' scores did not improve when self-assessment was used (see Appendices A through I).

Risk scores were also analyzed using ANCOVA and no relationship was found between the number correct for each gender, ethnicity, treatment (SA, NOSA) and risk.

Based on the results of the pilot study, a redesigned study was conducted, this time including ethnicity as a variable. Because of the small number of Native Americans and Blacks in the subject pool, only Hispanic and Non-Hispanic subjects were tested.

Another level was added to the independent variable Treatment (SA, NOSA). The added level may be described as a usability assessment (UA) group; subjects in this group were required to assess the usability of each test item.

Usability assessment was included as a control group to account for possible confounding behaviors. Subjects in the usability assessment groups performed the same type of motor movements and engaged in a similar type of reflective thinking process as subjects in the self-assessment groups. Instead of indicating a level of sureness for each answer, subjects indicated how useful they felt the information was. Usability assessment was also used to determine if making self-assessments about the sureness of answers improves performance, or if engaging in reflective thinking after answering test items improves performance.

Subjects were given the same sample SAT test as the SA and NOSA group, but marked their answers on a modified SA answer sheet (see Appendix J). Subjects first selected an answer and then assessed the usability of the information; there were five "useful" categories to choose from. They ranged from "Not Useful At All" to "Extremely Useful."

There are performance differences between males and females in multiple-choice testing. Self-assessment seems to improve performance for females by allowing them to express their level of sureness or unsureness in the correctness of their answers (facilitates risk) (Hassmen & Hunt, 1990). There are also performance differences between ethnicities in multiple-choice testing (Isonio, 1990).

By including ethnicity as a variable, and adding another level to the variable treatment, the current study, described here, was conducted with the hypotheses stated below:

Chapter 2

HYPOTHESES

It is hypothesized that females who self-assess will achieve a significantly higher score on the multiple-choice test than females who don't self assess. This difference may be explained by analyzing females' risk scores. Females who self-assess should have less conservative risk scores than females who don't self-assess.

Performance on the test depends not only on gender and treatment (SA, NOSA), but on ethnicity as well. It is hypothesized that Hispanics who self-assess will achieve higher test scores than Hispanics who don't self-assess.

Method

Subjects

Two hundred and forty undergraduate students from introductory psychology courses volunteered to serve as subjects.

Subjects were randomly assigned to 3 treatments, with the restriction that each treatment group would have an equal number of males and females and Hispanics and Non-Hispanics in it. As a result, 12 subgroups were

formed with 20 of each gender and ethnicity per group (see Table 3). Each subject received one credit hour of Psychology 201 for their participation in the study.

Table 3

Sample Sizes For Each Ethnicity, Gender and Treatment

SUBJECT	UA	SA	NOSA
Non-Hispanic Males	20	20	20
Non-Hispanic Females	20	20	20
Hispanic Males	20	20	20
Hispanic Females	20	20	20

Design/Instruments

The overall design may be described as a between-subjects, 2 X 2 X 3 factorial with the dependent variables being number of correct responses and risk score, and the independent variables being Gender: Male and Female; Ethnicity: Non-Hispanic and Hispanic; and Treatment: Usability Assessment, Self-Assessment, and No Self-Assessment. For the purpose of this experiment, the alpha level was set at 0.05.

Each subject was administered the fifty-item multiple-choice test developed by Hassmen and Hunt (1990)

(see Appendix B). The fifty items were extracted from different SAT tests; items chosen were determined to be as "gender equal" as possible. Evenly spaced throughout the test were ten mathematical questions; the remaining forty questions measured verbal ability. Each test question had five optional answers with only one being correct.

The NOSA groups marked their answers on the standard multiple-choice answer sheet (see Appendix C). The SA groups marked their answers on the "Multiple-Choice Self-Assessment Answer Sheet developed by Hunt (1990 Version) (see Appendix I). The UA groups marked their answers on the modified Multiple-Choice Self-Assessment Answer Sheet (see Appendix J).

All subjects were given the risk-taking questionnaire developed by Kogan and Wallach (1964) entitled "Choice Dilemmas Procedure: Opinion II Questionnaire" (see Appendix E).

The twelve-item test was administered after the SAT multiple-choice test. The test items represent choices between "risky and safe courses of action" (Kogan & Wallach, 1964). The instrument is semi-projective in nature. The subject is asked to give advice to different individuals in different situations. Kogan and Wallach (1964) assume "that an individual's advice to others

reflects his or her own regard for the desirability of success relative to the disutility of failure" (p.6).

There are six probability levels: 1 in 10, 3 in 10, 5 in 10, 7 in 10, 9 in 10, and subjects are given an additional choice NOT to take any risks, no matter what the probabilities. A ten is given for that response. The subject's choices are then summed and that becomes his or her risk score. The higher a subject's score, the more conservative he or she is considered to be. A subject's risk-taking score could range from 12 to 120. Subjects marked their choices directly onto the test itself.

Procedure

Subjects volunteered to participate in the experiment by signing their names on experimental sign-up sheets posted on the Psychology Department's bulletin board; ethnic group membership was based on self-identification. Sign-up sheets were posted by the experimenter every two weeks; subjects had their choice of test date. Each sign up sheet was divided into four cells: Non-Hispanic males and females and Hispanic males and females.

Test sessions were conducted until there were 20 subjects per subgroup. Test sessions were conducted every Tuesday afternoon at two o'clock; all three treatment groups were tested at each session.

After verifying attendance, subjects were informed that the purpose of the study was to examine different multiple-choice testing methods. Each subject was then given a folder which contained either a standard (NOSA) multiple-choice answer sheet, a self-assessment (SA) answer sheet, or a usability (UA) answer sheet. Each folder also contained written instructions on how to use the answer sheet in the folder (see Appendices H, K, and L), written instructions pertaining to the SAT test (see Appendix F), and a piece of plain bond paper to be used as "scratch" paper.

Subjects were asked to write their names, social security numbers, gender, age, GPA, and ethnicity in the appropriate spaces on the front of the folder. They were also instructed to put their names and social security numbers on their respective answer sheets.

Subjects were then given time to read the written instructions pertaining to the use of their particular answer sheets. No verbal instructions were given. Verbal instructions were then given concerning the actual

test itself (see Appendix F) and subjects were informed that each folder contained the same instructions in written form.

The tests were passed out and the subjects were given permission to begin. They were informed they had 45 minutes to complete the test.

Upon completion of the test, answer sheets and exams were put in the folders and verbal instructions were given for the risk-taking test (see Appendix G). Each subject was given a risk-taking test and given permission to begin. The risk taking test was not timed.

Results

Separate analyses were conducted on the performance measures: number of correct responses and risk score. A significance level of .05 was used. The means and variances for the number of correct responses for the various groups are provided in Table 4.

Results of Bartlett's test for homogeneity of variances performed on number of correct responses revealed that the variances among the twelve groups were not statistically different $\chi^2(11, N = 20) = 6.53, p > .05$ (see Appendix M).

An Analysis of Covariance (ANCOVA) was performed on the number of correct responses. GPA was used as the covariate to adjust for chance differences between the groups. The ANCOVA revealed a significant two-way interaction between ethnicity and gender $F(1, 239) = 4.75, p < .05$, and between ethnicity and treatment $F(2, 239) = 3.57, p < .05$. Effects of ethnicity were different depending on gender and treatment (see Figures 1 and 2) (see Appendix M for ANCOVA table).

Table 4

Means and Variances for Number of Correct Responses for Treatment, Ethnicity, and Gender Based on 20 Observations Per Group

Treatment	Ethnicity	Gender	Mean	Variance
Usability Assessment	Non-Hispanic	M	29.3	41.6
		F	30.2	44.5
	Hispanic	M	23.2	31.8
		F	18.6	35.9
Self-Assessment	Non-Hispanic	M	26.4	61.9
		F	26.6	34.0
	Hispanic	M	21.0	39.8
		F	19.7	32.4
No Self-Assessment	Non-Hispanic	M	27.3	34.5
		F	27.9	43.1
	Hispanic	M	26.2	68.9
		F	21.4	43.7

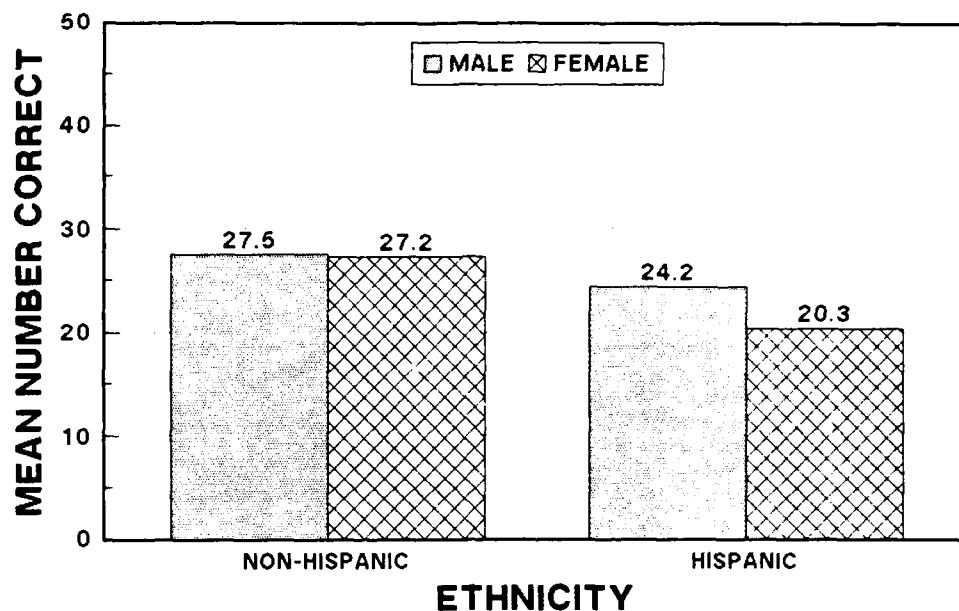


Figure 1. Mean number correct for Non-Hispanics and Hispanics by gender.

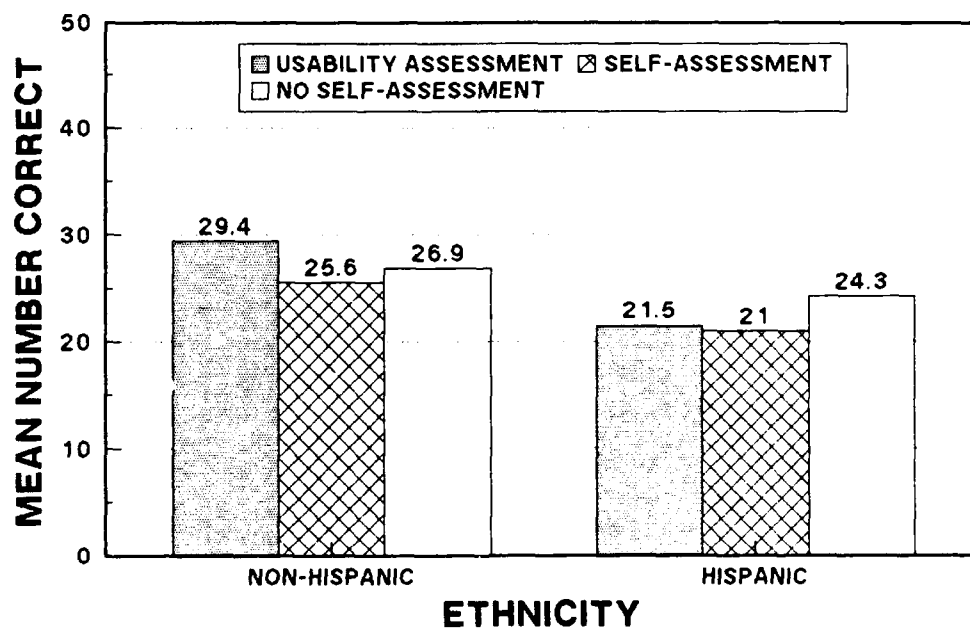


Figure 2. Mean number correct for Non-Hispanics and Hispanics by treatment.

Subsequently, means were compared using the protected Least Significant Difference (LSD) procedure to assist in interpreting both interactions. LSDs revealed the information contained in Tables 5 and 6.

Table 5

Ethnicity and Gender Mean Pairings From Protected Least Significant Difference Comparisons. Means With the Same Letter are not Significantly Different

Protected L.S.D. Group Mean Comparisons			
	A	Non-Hispanic Male	27.5
	A	Non-Hispanic Female	27.2
	A	Hispanic Male	24.2
B		Hispanic Female	20.3

Table 6

Ethnicity and Treatment Mean Pairings From Protected Least Significant Difference Comparisons. Means With the Same Letter are not Significantly Different

Protected L.S.D. Group Mean Comparisons			
		A	Non-Hispanic Usability Assessment 29.2
	B	A	Non-Hispanic No Self-Assessment 26.9
	B	A	Non-Hispanic Self-Assessment 25.6
C	B		Hispanic No Self-Assessment 24.3
C			Hispanic Usability Assessment 21.5
C			Hispanic Self-Assessment 21.0

Non-Hispanic males' scores, Non-Hispanic females' scores, and Hispanic males' scores did not differ statistically from each other. However, Hispanic females' scores were statistically lower than these three groups.

There were no significant differences in the means of the three treatments for each ethnicity.

Differences were found in treatment means between ethnicities. Non-Hispanics who tested with the usability assessment answer sheet scored significantly higher than Hispanics from all three treatment groups. Non-Hispanics who self-assessed, and those who were tested without self-assessment scored significantly higher than Hispanics who made usability assessments and Hispanics who self-assessed. There were no significant differences between the scores of Non-Hispanics who self-assessed, Non-Hispanics who did not self-assess, and Hispanics who did not self-assess.

Risk scores were collected from all subjects in each group. The means and variances for the risk scores for the various groups are provided in Table 7.

Bartlett's test for homogeneity of variance revealed that the risk score variances for each group were not equal, $\chi^2(11, N = 20) = 41.9, p < .001$. Subsequently, a

nonparametric procedure, the Kruskal-Wallis Test, was performed on the risk scores.

Table 7

Means and Variances for Risk Scores for Treatment, Ethnicity, and Gender Based on 20 Observations Per Group

Treatment	Ethnicity	Gender	Mean	Variance
Usability Assessment	Non-Hispanic	M	69.6	256.0
		F	67.5	401.3
	Hispanic	M	65.6	176.4
		F	70.3	239.2
Self-Assessment	Non-Hispanic	M	68.3	155.6
		F	66.1	164.7
	Hispanic	M	68.8	130.6
		F	66.5	246.3
No Self-Assessment	Non-Hispanic	M	62.9	109.2
		F	68.4	235.9
	Hispanic	M	65.0	894.6
		F	78.1	134.0

Results revealed that the mean risk score for female Hispanics tested without self-assessing was significantly higher ($\bar{M} = 78.1$) than the mean risk score for male Non-Hispanics tested without self-assessing ($\bar{M} = 62.9$), $\chi^2(11, N = 20) = 19.8, p < .05$.

While the risk scores for Hispanic females and Non-Hispanic males tested without self-assessing were significantly different from each other, neither one alone was different from the rest of the groups.

Chapter 3

DISCUSSION

Results of this study do not support the overall hypothesis that females, regardless of ethnicity, who engage in self-assessment during testing achieve a significantly higher score on the multiple-choice test than females who do not engage in self-assessment. The risk scores for these two groups were not significantly different; self-assessment did not improve females' scores. Additionally, self-assessment appeared to be detrimental for Hispanic males.

The findings concerning self-assessment are not consistent with results of Sams' (1986) study. She found that females who engaged in overt self-assessment responding while learning obtained a higher percentage of correct responses during learning trials and on a test than those who learned without self-assessment (Sams, 1986).

Hassmen and Hunts' (1990) self-assessment experiment showed significant main effects of gender and treatment. Hassmen and Hunt (1990) found female SA and female NOSA groups differed significantly $p < .01$; females who self-assessed performed significantly better than females who did not. Males' scores did not improve significantly.

Although the results of the pilot study, which preceded the current study, were not statistically significant ($p = .07$), the data suggested something of interest might be occurring as revealed by the three-way interaction of gender, ethnicity, and treatment. In that study, self-assessment appeared to have had a positive impact for Hispanic males and females. When self-assessment was used, significant differences between Hispanic and Non-Hispanic, and male and female scores disappeared.

In the current study, a significant interaction was found between ethnicity and gender. No significant differences were noted between the scores of Non-Hispanic males and females, and Hispanic males. However, these three groups scored significantly higher than Hispanic females.

According to Feingold (1988), cognitive gender differences are disappearing; the only exception to this trend is at the highest end of the mathematics-ability continuum, where the ratio of males outscoring females has remained constant over the years. Feingold's conclusions are based on a longitudinal review of gender differences on the Differential Aptitude Tests (DAT) and Preliminary Scholastic Aptitude Test/Scholastic Aptitude Test (PSAT/SAT). No explanation is given as to why the

change in cognitive differences has occurred.

Feingold's study did not address cognitive differences between ethnicities.

Feingold's predictions are not consistent with the results of the current study; the predictions seem to be relevant to the Non-Hispanic population only. Non-Hispanic females' scores did not differ from Non-Hispanic males' scores and Hispanic males' scores. However, Hispanic females' scores were significantly different from those three groups. A gender gap still exists for female Hispanics.

Mestre (1988) contends that Hispanic parents tend to encourage their daughters to focus on their future families rather than on educational endeavors. This parental stereotype may result in poorer test performance for Hispanic females.

A significant interaction was also found between ethnicity and treatment. For each ethnicity alone no statistically significant differences were found among the three treatments. Allowing test takers to indicate the level of their sureness in their answers by using the SA answer sheet, or to indicate the usability of the information contained in the test by using the UA answer sheet, did not appear to improve or degrade their scores when compared to the standard multiple-choice (NOSA)

answer sheet. Each ethnicity scored equally well on the test using the UA, SA and NOSA answer sheets.

However, there were significant differences between ethnicities and treatments. Non-Hispanics making usability assessments scored higher than Hispanics from all three treatment groups. The process of reflecting after each answer and assessing the usefulness of test items seemed to benefit Non-Hispanics. Non-Hispanics tested with and without self-assessing scored higher than Hispanics making usability and self-assessments. Non-Hispanics tested with and without self-assessing scored as well as Hispanics tested without self-assessing.

Hispanics' test performance is degraded compared to Non-Hispanics test performance when making self and usability assessments. Perhaps the time spent making assessments inhibits the performance (accuracy) of Hispanics when testing using these methods.

Llabre and Froman (1987) found that Hispanic examinees consistently spent more time than Non-Hispanic examinees on standard multiple-choice test items, had higher omission rates, and that imposing a time constraint seemed to penalize the Hispanic examinees.

In the current study, Hispanic examinees completed the test on time and omission rates were insignificant. However, Hispanics scored lower than Non-Hispanics when

tested with the usability and self-assessment answer sheets. That phenomenon was not noted when the NOSA answer sheet was used.

The data collected by the Opinion II Questionnaire (risk test) do not support the prediction that females who self-assessed would have higher risk-taking propensities than females who did not self-assess. The only differences noted in risk-taking were between female Hispanics and male Non-Hispanics tested using the NOSA answer sheet. Female Hispanics were found to be more conservative compared to male Non-Hispanics. Neither group differed significantly from the other treatment groups.

This current study was not an exact replication of Hassmen and Hunts' (1990) study, but was fairly close. The following experimental conditions were the same for both experiments: (a) the same 50 item test was used; (b) equal sample sizes were tested; (c) self-assessors and non-self-assessors were tested together; (d) subjects were tested in large classrooms with single desks; (e) each group was given verbal instructions concerning the test itself, and written instructions on how to use their respective answer sheets; (f) self-assessors were aware they could receive extra points for making correct self-

assessments; and (g) test dates and times were the same for all groups.

The major differences between the experiments were that a control group (Usability Assessment) was added to the current study, and each subject was asked to identify his or her ethnicity. Hassmen and Hunt did not collect data concerning ethnicity.

Also during the time that Hassmen and Hunt conducted their study, Hunt taught several undergraduate Psychology classes and occasionally tested Psychology 201 students using the self-assessment answer sheet. It may be that some of those students who were tested using those sheets also participated in the Hassmen and Hunt study.

The self-assessment process has been shown to be beneficial in the area of learning and testing (Hassmen & Hunt, 1990; Hunt, 1982, & Sams, 1986). Currently, similar self-assessment testing methods are being used in the Los Angeles School District. Results appear favorable.

Different results for this study may have been obtained had Psychology 201 students been more familiar with the SA answer sheet.

Results of this study show that:

1. Hispanic females scored significantly lower than Hispanic males and Non-Hispanic males and females on the multiple-choice test

2. Hispanics do not perform as well as Non-Hispanics when using usability and self-assessment answer sheets.

Further research is needed to investigate gender and ethnicity differences in test performance and, if possible, to determine what factors are responsible for such differences in performance. Research is also needed to determine the best possible testing methods to employ so that differences between Hispanic and Non-Hispanic test takers can be alleviated.

REFERENCES

- Aiken, L.R. (1979). Educational values of Anglo-American and Mexican American college students. The Journal of Psychology, 102, 317-21.
- Aiken, L.R. (1987). Testing with multiple-choice items. Journal of Research and Development In Education, 20, 44-58.
- Anderson, R.I. (1982). Computer based confidence testing alternatives to conventional computer based multiple-choice testing. Journal of Computer Based Instruction, 91, 1-19.
- Angoff, W.H. (1971). The College Board Admission Testing Programs: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board.
- Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. Journal of Educational Measurement, 28, 23-35.
- Bokhorst, F.D. (1986). Confidence-weighting and the validity of achievement tests. Psychological Reports, 59, 383-386.
- Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. Journal of educational Measurement, 27, 165-174.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Clark, M., & Grandy, J. (1984). Sex differences in the academic performance of scholastic aptitude test takers (College Board Report No. 84-8 ETS RR No. 84-43). New York: College Entrance Examination Board.
- Cruise, J., & Trusheim, D. (1988). The case against the SAT. Chicago and London: The University of Chicago Press.

- Duran, R.P. (1983). Hispanics' education and background: Predictors of college achievement. New York: College Entrance Examination Board.
- Duran, R.P. (1988). Bilinguals' logical reasoning aptitude: A construct validity study. In R. R. Cocking & J.P. Mestre (Eds.). Linguistic and cultural influences on learning mathematics. Hillsdale, NJ: Lawrence L. Baum Assoc.
- Echternacht, G.J. (1972). The use of confidence testing in objective tests. Educational Testing Services, 42(2), 217-236.
- Feingold, A. (1988). Cognitive gender differences are disappearing. American Psychologist, 43(2), 95-103.
- Glass, G.V., & Wiley, D.E. (1964). Formula scoring and test reliability. Journal of Educational Measurement, 1(1), 43-47.
- Goldman, S., & Duran, R. (1988). Answering questions from oceanography texts: Learner, task, and test characteristics. Discourse Processes, 11, 373-412.
- Goldman, R.D., & Newlin-Hewitt, B. (1975). An investigation of test bias for Mexican-American college students. Journal of Educational Measurement, 12, 187-196.
- Grebler, L., Moore, J.W., & Guzman, R.C. (1970). The family: Variations in time and space. In I.L. Duran & H.R. Bernard (Eds.). Introduction to Chicano Studies (pp. 309-331). New York, NY: Macmillan.
- Hassmen, P., & Hunt, D.P. (1990). Human self-assessment: A method to reduce gender bias in multiple-choice testing. Manuscript submitted for publication.
- Hoffman, B. (1962). The tyranny of testing. New York: Crowell-Collier.
- Hopkins, K.D., Hakstian, A.R., & Hopkins, B.R. (1973). Validity and reliability consequences of confidence weighting. Educational and Psychological Measurement, 33, 135-141.

- Hudgens, G.A., & Fatkin, L.T. (1985). Sex differences in risk-taking: Repeated sessions on a computer simulated task. Journal of Psychology, 119(3), 197-206.
- Hunt, D.P. (1982). Effects of human self-assessment responding on learning. Journal of Applied Psychology, 67, 75-82.
- Hunt, D.P. (1984). Human self-assessment and the implications for human training and performance. (Contract No. MDA 903-80-C-0276). Washington, DC: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Hunt, D.P. (1991). Self-assessment technology: Multiple-choice self-assessment testing. Human Performance Enhancement Inc., 1-12.
- Isonio, S. (1990). LAUSD student performance on the 1988-89 Scholastic Aptitude Test: Description and comparative analysis (Report No. 549). LAUSD, CA Program and Evaluation Branch. (ERIC Document Reproduction Service No. ED 331-9181)
- Jacobs, S.S. (1971). Correlates of unwanted confidence in responses to objective test items. Journal of Educational Measurement, 8, 15-19.
- Kogan, N., & Wallach, M.A. (1964). Risk-taking: A study in cognition and personality. New York: Holt, Rinehart & Winston.
- Llabre, M.M., & Froman, T.W. (1987). Allocation of time to test items: A study of ethnic differences. Journal of Experimental Education, 55, 137-40.
- Llabre, M.M., & Froman, T.W. (1988). Allocation of time and item performance in Hispanic and Anglo examinees (Final Report). University of Florida, Miami: Institute for Student Assessment and Evaluation.
- Lord, F.M. (1963). Formula scoring and validity. Educational and Psychological Measurement, 23, 663-672.
- Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, Calif.: Stanford University Press.

- MacCorquodale, P. (1988). Mexican-American women and mathematics: Participation, aspirations, and achievement. In R. Cocking & J. Mestre (Eds.), Linguistic and cultural influences on learning mathematics, (pp. 137-160). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mestre, J.P. (1988). The role of language comprehension in mathematics and problem solving mathematics and problem solving. In R. Cocking & J. Mestre (Eds.), Linguistic and cultural influences on learning mathematics, (pp. 201-220). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Palmer, M.T. (1990). Gender differences in cognitive performance in response to different levels of anxiety: An evolutionary approach. Unpublished master's thesis. New Mexico State University, Las Cruces.
- Pascale, P.J. (1974). Changing initial answers on multiple-choice achievement tests. Measurement and Evaluation in Guidance, 6, 236-238.
- Plax, T.G., & Rosenfeld, L.B. (1976). Correlates of Risky Decision Making. Journal of Personality Assessment, 40(4), 413-418.
- Rippey, R.M., & Voytovich, A.E. (1985). Anomalous responses on confidence scored tests. Evaluation & The Health Profession, 8(1), 109-120.
- Rosser, P. (1989). The SAT Gender Gap: Identifying the Causes. Washington, DC: Center for Women Policy Studies.
- Sams, M.R. (1986). Effects of testing methods and self-assessment responding on observers and performers. Unpublished master's thesis, New Mexico State University, Las Cruces.
- Sams, M.R. (1989). Effects of observational assessments and patterns of success-failure on self-confidence. Unpublished dissertation. New Mexico State University, Las Cruces.
- Schwartz, A.J. (1971). A comparative study of values and achievement: Mexican-American and Anglo youth. Sociology of Education, 44, 438-462.

- Schmitt, A.P., & Dorans, N.J. (Eds.), (1987).
Differential item functioning on the Scholastic Aptitude Test. Princeton, NJ: Educational Testing Service.
- Schmitt, A.P. (1988). Language and cultural characteristics that explain differential item functioning for Hispanic examinees on the Scholastic Aptitude Test. Journal of Educational Measurement, 25, 1-13.
- Shuford, E.H., Albert, A., & Massengill, H.E. (1966). Admissible probability measurement procedures. Psychometrika, 31, 125-145.
- Skinner, N.F. (1983). Switching answers on multiple-choice questions: Shrewdness or shibboleth? Teaching of Psychology, 10, 220-222.
- Slakter, M.J. (1968). The penalty for not guessing. Journal of Educational Measurement, 52, 141-144.
- Slovic, P. (1964). Assessment of risk-taking behavior. Psychological Bulletin, 61(3), 220-233.
- Swineford, F. (1938). The measurement of a personality trait. Journal of Educational Psychology, 29, 295-300.
- Temp, G. (1971). Validity of the SAT for Blacks and Whites in thirteen integrated institutions. Journal of Educational Measurement, 8, 245-251.
- Wood, R. (1976). Inhibiting blind guessing: The effect of instructions. Journal of Educational Measurement, 13, 297-307.
- Younkin, W. (1986). Speededness as a source of test bias on the College Level Academic Skills Test. Unpublished doctoral dissertation, University of Miami.

APPENDIX A

Pilot Study

Pilot Study

The pilot study, described here, was conducted to select suitable methods, procedures, and testing materials so that an improved study could be performed to determine whether: 1) using self-assessment during testing improves a test taker's score, i.e., the number correct, 2) females who self-assess achieve a higher number correct than females who don't self-assess, and 3) this hypothesized difference, if it exists, can be interpreted using the subject's risk propensity score.

Method

Subjects

One-hundred thirteen undergraduate students who were enrolled in Psychology 201 at New Mexico State University served as subjects. Initially 120 volunteered; 7 failed to show. Sixty-one were female and 52 were male (see Table 1 for information regarding ethnicity). Each subject received one credit hour for their participation.

Subjects were randomly assigned to a control group (standard multiple-choice test answer sheets were used), or an experimental group (self-assessment answer sheets were used). Random assignment was accomplished by

posting sign-up sheets which reflected different test dates. Testing began on 20 January and ended on 21 February 1992. Testing was conducted every Monday and Friday at two o'clock in the afternoon. Order of treatments was counterbalanced. For example, on the first Monday, subjects were administered the test using the self-assessment answer sheet, and those subjects who participated on Friday were tested using the standard multiple-choice answer sheet. The next week the order was switched.

Appendix Table A1

**Sample Sizes for Each Ethnicity, Gender, and Treatment
No Self-Assessment-NOSA, Self-Assessment-SA**

ETHNICITY	GENDER	NOSA	SA
Non-Hispanic	M	15	15
	F	17	17
Hispanic	M	8	8
	F	10	8
Native American	M	2	2
	F	3	5
Black	M	0	2
	F	1	0

Instruments

A 50-item multiple-choice test developed by Hassmen and Hunt (1990) was used (see Appendix B). The 50 items were extracted from different SAT tests; items chosen were determined to be as "gender equal" as possible. Evenly spaced throughout the test were ten mathematical questions; the remaining 40 questions measured verbal ability. Each test question had five alternative answers with only one being correct. The control groups answered the questions using the standard multiple-choice answer sheet (see Appendix C). After determining what they thought was the correct answer they marked the corresponding "bubble." The control group consisted of males and females; they will be referred to as Male NOSA and Female NOSA.

The experimental groups answered the same questions on a different multiple-choice answer sheet entitled, the "Multiple-Choice Self-Assessment Answer Sheet" (see Appendix D) developed by Hunt (1983). These subjects were instructed to answer each question by marking the appropriate "bubble" and then to immediately assess the correctness of that answer by marking one of five self-assessments ranging from "Almost a Guess" to "Almost Certain." The males and females in the experimental group will be referred to as Male-SA and Female-SA.

All subjects were given a risk-taking questionnaire developed by Kogan and Wallach (1964) entitled "Choice Dilemmas Procedure: Opinion II Questionnaire" (see Appendix E). The 12-item test was administered after the SAT multiple-choice test. The test items represent choices between "risky and safe courses of action" (Kogan & Wallach, 1964).

Kogan and Wallach (1964) assert that "A subject's selection of the probability level for the risky alternative's success that would make it sufficiently attractive to be chosen thus reflects the deterrence of failure for him in a particular decision area" (p.6).

The instrument is semi-projective in nature. The subject is asked to give advice to different individuals in different situations. Kogan and Wallach (1964) assume that an individual's advice to others reflects his own regard for the desirability of success relative to the disutility of failure.

There are six probability levels: 1 in 10, 3 in 10, 5 in 10, 7 in 10, 9 in 10, and subjects are given an additional choice NOT to take any risks, no matter what the probabilities. A ten is given for that response. The subject's choices are then summed and that becomes his or her risk score. The higher a subject's score, the

more conservative he or she is considered to be. A subject's risk-taking score could range from 12 to 120.

Subjects marked their choices directly onto the test itself.

The overall design of this experiment may be described as a between subjects, 2 X 2 factorial, with the independent variables being: Self-Assessment-SA and No Self-Assessment-NOSA, and Gender-Male (M) and Female (F). The dependent variable is test performance (accuracy) measured in number correct. The risk propensity score is merely a tool used to interpret the hypothesized difference in scores.

For the purpose of this pilot study only, the alpha level was set at .10.

Procedure

There were ten test sessions; an equal number of subjects was not tested at each session because some scheduled subjects failed to appear. After verifying attendance, subjects were given an answer sheet and asked to put their name, gender, grade point average (GPA), ethnicity, and age at the top of the sheet. GPA, ethnicity, and age were requested from the subjects to account for possible variance in scores.

Verbal instructions were given on how to use the answer sheets. These instructions differed slightly (see Appendices C and D) depending on the answer sheet being used. Control groups and experimental groups were tested separately whereas Hassmen and Hunt (1990) tested control and experimental groups together. They also tested more subjects per session ($n=40$). Hassmen and Hunt (1990) gave written instructions on how to use the answer sheets.

Additional verbal instructions were given concerning the actual test itself (see Appendix F). The tests were passed out and the subjects were given permission to begin. They were informed they had 45 minutes to complete the test.

Upon completion of the test, answer sheets and tests were collected and the instructions were read for the risk-taking test (see Appendix G). Each subject was given a risk-taking test and given permission to begin. The risk-taking test was not timed.

Results

An Analysis of Covariance (ANCOVA) revealed a significant three-way interaction among gender, self-assessment, and ethnicity, ($p = 0.07$). The covariates were age and grade point average (GPA). Effects of self-

assessment were different depending on gender and ethnicity. Subsequently, multiple comparisons among means were conducted using the protected Least Significant Difference (LSD) test to assist in interpreting the 3 way interaction. The significance level of 0.07 was used for the LSD procedure (M. Ortiz, personal communication, 28 July 1992). LSDs revealed the following information:

When females were tested without self-assessing (NOSA), no statistical differences were noted between the scores of Non-Hispanics and Native Americans; they performed equally well on the multiple-choice test (note the small n for Native Americans). However, Hispanics scored significantly lower than Non-Hispanics. Hispanics' scores did not differ statistically from Native Americans' scores (see Table 2).

When females were tested using self-assessment, differences between Hispanics' and Non-Hispanics' scores disappeared. Native Americans performed significantly lower than both Non-Hispanics and Hispanics.

When males were tested without self-assessing (NOSA), Non-Hispanics scored significantly higher than Hispanics and Native Americans. Hispanics' scores did not differ statistically from Native American scores.

When males were tested using self-assessment, no differences were found among the three ethnicities.

Hispanic females (NOSA) scored significantly lower than Non-Hispanic males (NSA), but when both were tested using self-assessment those differences disappeared.

When Native American females self-assessed, they achieved much lower scores than Non-Hispanic males (NOSA) and both Non-Hispanic and Hispanic males using self-assessment.

Appendix Table A2

Means for Number of Correct Responses, and Sample Sizes for Ethnicity, Gender, and Treatment
No Self-Assessment-NOSA, Self-Assessment-SA

ETHNICITY	GENDER	NOSA (\bar{x} , n)	SA (\bar{x} , n)
Non-Hispanic	M	28.8, 15	24.0, 15
	F	27.2, 17	25.0, 17
Hispanic	M	23.1, 8	24.3, 8
	F	20.8, 10	24.2, 10
Native American	M	19.5, 2	26.5, 2
	F	26.6, 3	17.4, 5

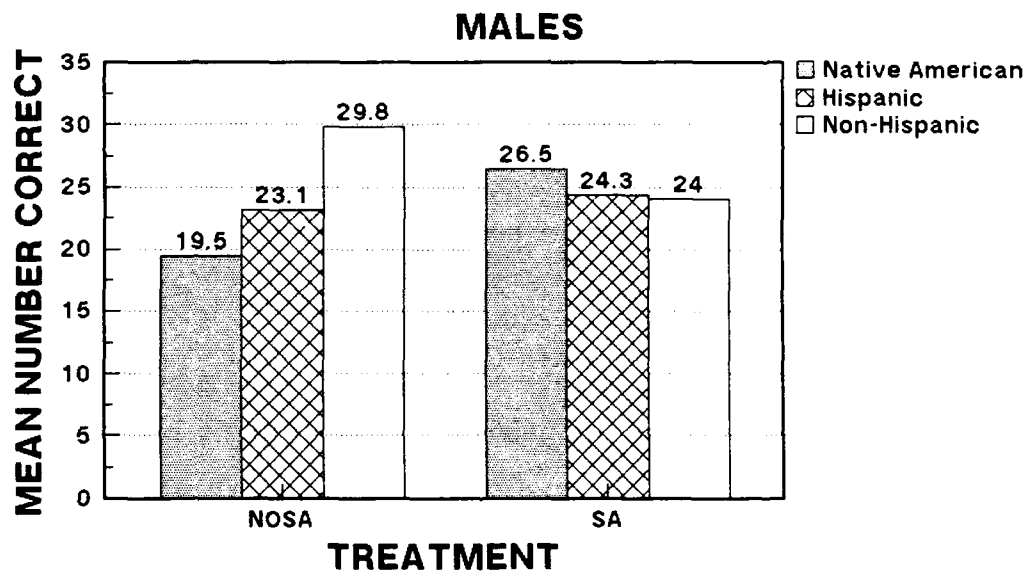
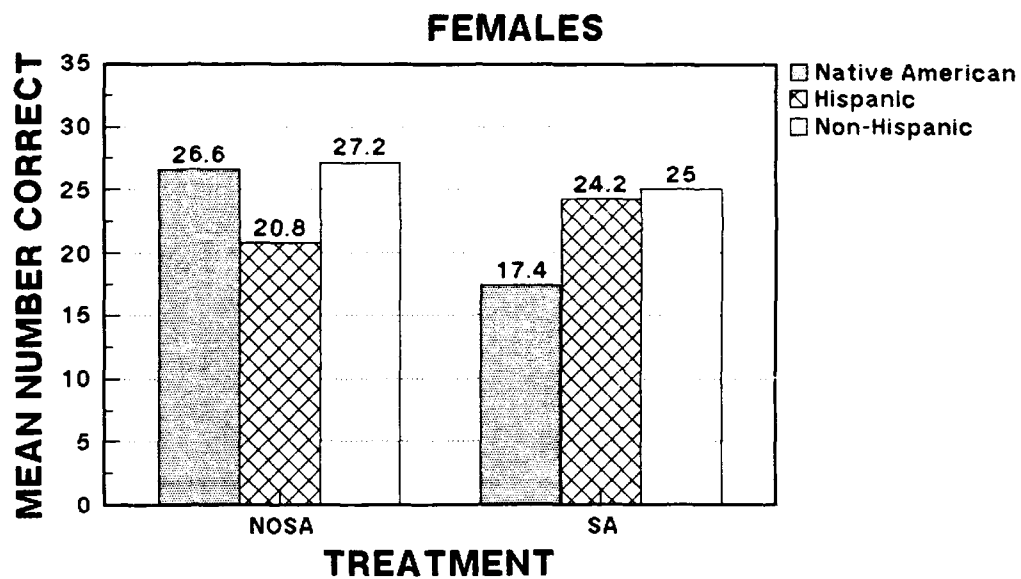
Risk scores were collected from all subjects and were also analyzed using ANCOVA. GPA and age were the

covariates. No relationships were found between the number correct for each gender, ethnicity, treatment (NOSA, SA), and risk score (see Table 3 for mean risk scores).

Appendix Table A3

**Means for Risk Scores, and Sample Sizes
for Ethnicity, Gender, and Treatment
No Self-Assessment-NOSA, Self-Assessment-SA**

		Risk Scores per Treatment	
ETHNICITY	GENDER	NOSA (\bar{x} , n)	SA (\bar{x} , n)
Non-Hispanic	M	63.9, 15	76.0, 15
	F	75.1, 17	69.4, 17
Hispanic	M	77.0, 8	73.2, 8
	F	76.5, 10	71.2, 8
Native American	M	76.0, 2	78.0, 2
	F	71.3, 3	72.6, 5



Appendix Figure A 1. Mean number correct for each treatment for Native Americans, Hispanics, and Non-Hispanics by gender.

Discussion

Results of this study do not support the overall hypothesis that females who self-assess achieve a significantly higher score on the multiple-choice test than females who do not engage in self-assessment. There was no significant difference between the two groups' risk propensity scores. However, when taking ethnicity into account, it appears that self-assessment may be beneficial for Hispanic females and males, and neutral to Native American females and Non-Hispanic males.

These findings are not consistent with Sams (1986), who found that females' performance was positively affected when self-assessment was used, and Hassmen and Hunts' (1990) results which showed significant main effects of gender and treatment. Hassmen and Hunt (1990) found female SA and female NOSA groups differed significantly, ($p < .01$); females who self-assessed performed significantly better than females who did not. Small sample sizes for Hispanics and Native Americans may be a reason for the inconsistent findings; therefore the interaction should be cautiously viewed.

The significant three-way interaction of gender, self-assessment, and ethnicity had a probability of error

equal to 0.07. Of course, this alpha level is higher than the more commonly used .05 level, but suggests that something of interest might be occurring.

Analyzing the data further using the protected Least Significant Difference procedure revealed that ethnicity played a major part in the interaction. For example, self-assessment appears to have had a positive impact for Hispanic females and Hispanic and Native American males. When self-assessment is used significant differences between Hispanic and Non-Hispanic, and male and female scores disappear.

It may be beneficial to conduct this study again to determine if ethnicity, gender, and self-assessment interact. Unfortunately, there are not enough Native Americans or Blacks available as subjects to pursue differences between their scores and the scores of the Non-Hispanics and Hispanics.

It is worth noting that this pilot study was not an exact replication of Hassmen and Hunts' (1990) study. The differences in the results of this experiment compared to Hassmen and Hunts' may be due to different experimental conditions and sample sizes. For example, Hassmen and Hunt (1990) tested the same number of subjects per session and more subjects per session ($n=40$). Because they tested more subjects at one time,

they were able to administer the test in a much larger classroom. Each subject was assigned to an individual desk. Due to space limitations, subjects who participated in the pilot study had to sit right next to each other at the same table. These space limitations may have influenced the subjects' performance.

Hassmen and Hunt (1990) also tested self-assessors and non-self-assessors together. Each group was given written instructions on how to use the answer sheets; no verbal instructions were given. Subjects who self-assessed were aware that they would receive extra points if they were sure of their answers.

They collected no data concerning ethnicity. It has been shown in this pilot study that ethnicity may be a major factor that one must consider in analyzing the data.

Considering the results of this pilot study, the following changes will be implemented in the proposed research and may better serve to determine the effects of self-assessment responding:

1. Fewer sessions will be conducted. More subjects will be tested per session. An equal number of males and females should be tested together. Also an equal number of Hispanics and Non-Hispanics should be tested each session.

2. A control group entitled, Usability Assessment group should be added to the design. This group would be required to assess how useful they think the information contained in the test is to them.
3. Verbal and written instructions should be given for the multiple-choice test, and only written instructions for the answer sheets. This may provide subjects with further clarification of what is expected of them.
4. More detailed instructions should be given to those subjects who self-assess. For example, they should know that they can earn extra points for being sure and correct (+50) compared to sure and wrong (-60). These improved instructions may be an incentive for subjects to do their best (see Appendix H). An updated version of the self-assessment answer sheet has been developed by Hunt (1990) (see Appendix I). This answer sheet is basically the same as the answer sheet developed by Hunt in 1983. Major changes include the condensing of self-assessment instructions and the rewording of the five alternatives. The five alternatives have been changed from Almost a

Guess, Probable Guess, Neutral, Fairly Certain, and Almost Certain to Not Sure At All, Very Unsure, Somewhat Sure, Very Sure, and Extremely Sure.

5. Day of testing may also be a factor to consider. Instead of testing on Mondays and Fridays, testing will be limited to the middle of the week, if possible.

There is gender bias associated with the Scholastic Aptitude Test. Using a multiple-choice test similar to the SAT, Hassmen and Hunt (1990) showed that when females were allowed to self-assess their scores improved significantly. These findings suggest that something about the self-assessment process seems to allow females to take risks by expressing the sureness or unsureness of their answers. Therefore, it is important to get a "risk score" after testing to see if there is a relationship between self-assessment and risk-taking. It is important to conduct a redesigned study, this time including ethnicity as an additional variable and incorporating the above mentioned changes.

APPENDIX B

Sample SAT Multiple-Choice Test (50 Items)

And Answer Key

1. CONVOKE:
 - (A) dissuade
 - (B) disperse
 - (C) reassure
 - (D) pacify
 - (E) diverge

2. NOSE : HEAD::
 - (A) hand : arm
 - (B) foot : toe
 - (C) eye : lid
 - (D) wrist : finger
 - (E) teeth : gums

3. In a family of five, the heights of the members are 5 feet 1 inch, 5 feet 7 inches, 5 feet 2 inches, 5 feet, and 4 feet 7 inches. The average height is
 - (A) 4 feet 4 inches
 - (B) 5 feet
 - (C) 5 feet 2 inch
 - (D) 5 feet 2 inches
 - (E) 5 feet 3 inches

4. FALLACIOUS:
 - (A) agreeable
 - (B) material
 - (C) verifiable
 - (D) exacting
 - (E) primary

5. WHEAT : GRAIN::
 - (A) cow : beef
 - (B) orange : citrus
 - (C) carrot : vegetable
 - (D) coconut : palm
 - (E) hamburger : steak

6. BELLICOSE:
(A) terse
(B) bleak
(C) inadequate
(D) pacific
(E) pliable
7. COTTAGE : CASTLE::
(A) house : apartment
(B) puppy : dog
(C) lot : acreage
(D) man : family
(E) poet : gentleman
8. $0.2 \times 0.02 \times 0.002 =$
(A) .08
(B) .008
(C) .0008
(D) .00008
(E) .000008
9. ABERRANT:
(A) distinguished
(B) proper
(C) seemly
(D) mindful
(E) calm
10. OLD : ANTIQUE::
(A) new : modern
(B) cheap : expensive
(C) useless : useful
(D) wanted : needed
(E) rich : valuable

11. AFFINITY:
(A) disrespect
(B) unfamiliarity
(C) antagonism
(D) distance
(E) ineptitude
12. DIGRESS : RAMBLE::
(A) muffle : stifle
(B) rust : steel
(C) introduce : conclude
(D) rest : stir
(E) find : explain
13. If the average weight of boys who are John's age and height is 105 lbs., and if John weighs 110% of the average, then how many pounds does John weight?
(A) 110
(B) 110.5
(C) 112
(D) 114.5
(E) 115.5
14. MOTIVE:
(A) vapid
(B) weak
(C) futile
(D) irrelevant
(E) inert
15. THROAT : SWALLOW::
(A) teeth : chew
(B) eyelid : wink
(C) nose : point
(D) ear : involve
(E) mouth : clamor

16. ELUSIVE:
(A) pragmatic
(B) constant
(C) decisive
(D) plodding
(E) sober
17. GARNET : RED::
(A) pearl : round
(B) diamond : solid
(C) emerald : green
(D) ivory : living
(E) silver : monetary
18. On a house plan on which 2 inches represents 5 feet, the length of a room measures 7.5 inches. The actual length of the room in feet is
(A) 12.5
(B) 15.75
(C) 17.5
(D) 18.75
(E) 19.25
19. RELENT:
(A) digress
(B) evade
(C) conclude
(D) encourage
(E) persevere
20. TRAVEL : JOURNEY::
(A) hop : stumble
(B) crawl : run
(C) lift : plane
(D) plan : itinerary
(E) walk : hike

21. CONSIDERATE
(A) instinctive
(B) vapid
(C) thoughtless
(D) noisy
(E) aloof
22. COTTON : SOFT::
(A) wool : warm
(B) iron: hard
(C) nylon : strong
(D) wood : polished
(E) silk : expensive
23. If five triangles are constructed having sides of the lengths indicated below, the triangle that will not be a right triangle is
(A) 5, 12, 13
(B) 3, 4, 5
(C) 8, 15, 17
(D) 9, 40, 41
(E) 12, 15, 18
24. LENIENT:
(A) intolerant
(B) punctual
(C) committed
(D) energetic
(E) inspired
25. YEAR : CENTURY::
(A) inch: yard
(B) mile : speed
(C) week : month
(D) cent : dollar
(E) day : year

26. RESTITUTION:
(A) inflation
(B) cataclysm
(C) deprivation
(D) benediction
(E) podium
27. CRACK : SMASH::
(A) merge : break
(B) run : hover
(C) s\whisper : scream
(D) play : work
(E) tatttle : tell
28. It costs \$1.30 a square foot to lay linoleum.
To lay 20 square yards of linoleum will cost
(A) \$47.50
(B) 49.80
(C) 150.95
(D) 249.00
(E) 234.00
29. CHIMERICAL:
(A) nimble
(B) realistic
(C) powerful
(D) underrated
(E) remarkable
30. MIDGET : SHORT::
(A) clown : fat
(B) actress : beautiful
(C) athlete : tall
(D) giant : big
(E) man : strong

31. INNOVATE:
(A) buy
(B) sell
(C) own
(D) copy
(E) choose
32. SPECTATOR : SPORT::
(A) jury : trial
(B) witness : crime
(C) soloist : music
(D) support : team
(E) fan : player
33. The total saving in purchasing 30 13-cent lollipops for a class party at a reduced rate of \$1.38 per dozen is
(A) \$.35
(B) \$.38
(C) \$.40
(D) \$.45
(E) \$.50
34. EULOGIZE:
(A) honor
(B) ignore
(C) defend
(D) berate
(E) heal
35. WALK : AMBLE::
(A) work : tinker
(B) play : rest
(C) run : jump
(D) fast : slow
(E) go: come

36. DOWNFALL:
(A) harm
(B) hazard
(C) weakness
(D) success
(E) quiet
37. TEA : LIQUID::
(A) potato : root
(B) corn : vegetable
(C) meat : food
(D) bread : solid
(E) coffee : cream
38. A gallon of water is equal to 231 cubic inches. How many gallons of water are needed to fill a fish tank that measures 11" high, 14" long, and 9" wide?
(A) 6
(B) 8
(C) 9
(D) 14
(E) 16
39. TURGID:
(A) dusty
(B) muddy
(C) rolling
(D) deflated
(E) tense
40. HAMMER : TOOL::
(A) tire : wheel
(B) wagon : vehicle
(C) nail : screw
(D) stick : drum
(E) saw : wood

41. IGNOMINY:
 (A) fame
 (B) isolation
 (C) misfortune
 (D) sorrow
 (E) stupidity
42. CLAP : THUNDER::
 (A) crowd : roar
 (B) hand : voice
 (C) bullet : cannon
 (D) scream : yell
 (E) bolt : lightning
43. A college graduate goes to work for \$x per week. After several months the company gives all the employees a 10% pay cut. A few months later the company gives all the employees a 10% raise. What is the college graduate's new salary?
 (A) .90 \$x
 (B) .99 \$x
 (C) \$x
 (D) 1.01 \$x
 (E) 1.11 \$x
44. DISPARAGE:
 (A) applaud
 (B) degrade
 (C) erase
 (D) reform
 (E) scatter
45. SPANK : PUNISH::
 (A) hit : beat
 (B) praise : reward
 (C) smile : flirt
 (D) wound : infect
 (E) act : require

46. OPULENT:
(A) fearful
(B) free
(C) oversized
(D) trustful
(E) impoverished
47. PROGRAM : COMPUTER::
(A) student : book
(B) conference : meeting
(C) recipe : cook
(D) index : book
(E) picture : photograph
48. What is the net amount of a bill of \$428.00 after a discount of 6% has been allowed?
(A) \$432.62
(B) \$430.88
(C) \$414.85
(D) \$412.19
(E) \$402.32
49. DEVIOUS:
(A) candid
(B) clever
(C) bright
(D) bitter
(E) vain
50. AWL : PUNCTURE::
(A) tire : flat
(B) cleaver : cut
(C) plane : area
(D) throttle : gas
(E) axle : wheel

ANSWER-KEY

- | | |
|-------|-------|
| 1. B | 26. C |
| 2. A | 27. C |
| 3. C | 28. E |
| 4. C | 29. B |
| 5. C | 30. D |
| 6. D | 31. D |
| 7. C | 32. B |
| 8. E | 33. D |
| 9. B | 34. D |
| 10. A | 35. A |
| 11. C | 36. D |
| 12. A | 37. D |
| 13. E | 38. A |
| 14. E | 39. D |
| 15. A | 40. B |
| 16. C | 41. A |
| 17. C | 42. E |
| 18. D | 43. B |
| 19. E | 44. A |
| 20. E | 45. B |
| 21. C | 46. E |
| 22. B | 47. C |
| 23. E | 48. E |
| 24. A | 49. A |
| 25. D | 50. B |

APPENDIX C

Standard Multiple-Choice Answer Sheet

NEW MEXICO STATE UNIVERSITY TEST SCORE SHEET

WRITE IN SOCIAL SECURITY
NUMBER AND BLACKEN IN
CIRCLE BELOW CORRESPONDING
TO NUMBER ENTERED

DIRECTIONS:

1. USE BLACK LEAD PENCIL ONLY (#2 OR SOFTER).
2. ONLY ONE RESPONSE PER QUESTION ALLOWED.
3. MAKE NO STRAY MARKS ON THIS SHEET.
4. ERASE CLEAN ANY MARK YOU WISH TO CHANGE.
5. DO NOT FOLD OR STAPLE THIS SHEET.

DO NOT MARK
OR WRITE HERE
FOR COMPUTER CENTER
USE ONLY

• • •

NAME _____ COURSE _____ SECTION _____

INSTRUCTOR _____ TEST _____ VERSION _____

DATE _____

SCORING ZONE - BLACKEN IN CIRCLE COMPLETELY CORRESPONDING TO YOUR CHOICE - ONLY ONE CHOICE PER QUESTION

1	T F	24	T F	47	T F	70	T F
2	T F	25	T F	48	T F	71	T F
3	T F	26	T F	49	T F	72	T F
4	T F	27	T F	50	T F	73	T F
5	T F	28	T F	51	T F	74	T F
6	T F	29	T F	52	T F	75	T F
7	T F	30	T F	53	T F	76	T F
8	T F	31	T F	54	T F	77	T F
9	T F	32	T F	55	T F	78	T F
10	T F	33	T F	56	T F	79	T F
11	T F	34	T F	57	T F	80	T F
12	T F	35	T F	58	T F	81	T F
13	T F	36	T F	59	T F	82	T F
14	T F	37	T F	60	T F	83	T F
15	T F	38	T F	61	T F	84	T F
16	T F	39	T F	62	T F	85	T F
17	T F	40	T F	63	T F	86	T F
18	T F	41	T F	64	T F	87	T F
19	T F	42	T F	65	T F	88	T F
20	T F	43	T F	66	T F	89	T F
21	T F	44	T F	67	T F	90	T F
22	T F	45	T F	68	T F	91	T F
23	T F	46	T F	69	T F	92	T F

T F 93	T F 116	T F 139	T F 162
T F 94	T F 117	T F 140	T F 163
T F 95	T F 118	T F 141	T F 164
T F 96	T F 119	T F 142	T F 165
T F 97	T F 120	T F 143	T F 166
T F 98	T F 121	T F 144	T F 167
T F 99	T F 122	T F 145	T F 168
T F 100	T F 123	T F 146	T F 169
T F 101	T F 124	T F 147	T F 170
T F 102	T F 125	T F 148	T F 171
T F 103	T F 126	T F 149	T F 172
T F 104	T F 127	T F 150	T F 173
T F 105	T F 128	T F 151	T F 174
T F 106	T F 129	T F 152	T F 175
T F 107	T F 130	T F 153	T F 176
T F 108	T F 131	T F 154	T F 177
T F 109	T F 132	T F 155	T F 178
T F 110	T F 133	T F 156	T F 179
T F 111	T F 134	T F 157	T F 180
T F 112	T F 135	T F 158	T F 181
T F 113	T F 136	T F 159	T F 182
T F 114	T F 137	T F 160	T F 183
T F 115	T F 138	T F 161	T F 184

APPENDIX D

Multiple-Choice Self-Assessment Answer Sheet

(1983 VERSION)

MULTIPLE CHOICE SELF-ASSESSMENT ANSWER SHEET

COPYRIGHT 1983 DARWIN HUNT - ALL RIGHTS RESERVED

DIRECTIONS:

1. FOR EACH QUESTION PLEASE SELECT THE BEST ANSWER AMONG THE GIVEN ALTERNATIVES USING BLACK LEAD PENCIL ONLY. #2 OR SOFTER COMPLETELY BLACKEN IN THE CIRCLE CORRESPONDING TO YOUR SELECTED ANSWER. ONLY ONE ANSWER PER QUESTION ALLOWED.
2. IMMEDIATELY AFTER YOU HAVE MARKED YOUR ANSWER PLEASE INDICATE HOW SURE YOU ARE THAT YOUR ANSWER IS CORRECT BY BLACKENING IN ONE OF THE FIVE LEVELS WHICH IS MOST REPRESENTATIVE OF YOUR CONFIDENCE IN THE CORRECTNESS OF YOUR ANSWER. ONLY ONE CONFIDENCE RESPONSE IS ALLOWED.
3. THE POINTS YOU RECEIVE ON EACH ITEM ARE BASED ON BOTH THE CORRECTNESS OF THE ANSWER AND THE ACCURACY OF YOUR SELF-ASSESSMENT INDICATING HOW SURE YOU ARE AS ACCURATELY AS YOU CAN WILL PRODUCE THE HIGHEST SCORE.
4. DO NOT FOLD OR STAPLE THIS SHEET. MAKE NO STRAY MARKS AND ERASE CLEAN ANY MARK YOU WISH TO CHANGE.

EXAMPLE: T F ANSWER HOW SURE ARE YOU?
 A B C ☒ D E F G H I J ☐ 0 1 2 ☒ 3

YOUR ANSWER IS ☐ D AND YOU ARE FAIRLY CERTAIN ☒ 3
 THAT YOUR ANSWER IS CORRECT

WRITE IN SOCIAL SECURITY NUMBER IN
 SPACE PROVIDED. BLACKEN IN CIRCLE BELOW
 CORRESPONDING TO NUMBER ENTERED

0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

DO NOT MARK IN THIS SPACE - FOR COMPUTER
 CENTER USE ONLY

NAME _____ DATE _____

COURSE TITLE _____ INSTRUCTOR _____

		HOW SURE ARE YOU?							HOW SURE ARE YOU?				
		ALMOST GUESS	PROBABLE GUESS	NEUTRAL	FAIRLY CERTAIN	ALMOST CERTAIN			ALMOST GUESS	PROBABLE GUESS	NEUTRAL	FAIRLY CERTAIN	ALMOST CERTAIN
ANSWER							ANSWER						
T F	1	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	2	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	3	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	4	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	5	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	6	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	7	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	8	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	9	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	10	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	11	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	12	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	13	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	14	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	15	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	16	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	17	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	18	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	19	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	20	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	21	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	22	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	23	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	24	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	25	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	26	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	27	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	28	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	29	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	
T F	30	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> 0	1	2	3	4	5	

DO NOT MARK
OR WRITE
IN THIS SPACE

		HOW SURE ARE YOU?							HOW SURE ARE YOU?				
ANSWER		ALMOST SURE	PROBABLE GUESS	NEUTRAL	FAIRLY CERTAIN	ALMOST CERTAIN	ANSWER		ALMOST SURE	PROBABLE GUESS	NEUTRAL	FAIRLY CERTAIN	ALMOST CERTAIN
31	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						46	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
32	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						47	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
33	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						48	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
34	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						49	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
35	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						50	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
36	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						51	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
37	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						52	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
38	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						53	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
39	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						54	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
40	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						55	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
41	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						56	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
42	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						57	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
43	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						58	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
44	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						59	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					
45	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z						60	T F A B C D E F G H I J K L M N O P Q R S T U V W X Y Z					

APPENDIX E

Risk-Taking Test

(12 Items: Developed by Kogan and Wallach 1964)

RISK-TAKING TEST

1. Mr. A, an electrical engineer, who is married and has one child, has been working for a large electronics corporation since graduating from college five years ago. He is assured of a lifetime job with a modest, though adequate, salary, and liberal pension benefits upon retirement. On the other hand, it is very unlikely that his salary will increase much before he retires. While attending a convention, Mr. A is offered a job with a small, newly founded company which has a highly uncertain future. The new job would pay more to start and would offer the possibility of a share in the ownership if the company survived the competition of the larger firms.

Imagine that you are advising Mr. A. Listed below are several probabilities or odds of the new company's proving financially sound.

Please check the lowest probability that you would consider acceptable to make it worthwhile for Mr. A to take the new job.

- ☐ A. The chances are 1 in 10 that the company will prove financially sound.
 - ☐ B. The chances are 3 in 10 that the company will prove financially sound.
 - ☐ C. The chances are 5 in 10 that the company will prove financially sound.
 - ☐ D. The chances are 7 in 10 that the company will prove financially sound.
 - ☐ E. The chances are 9 in 10 that the company will prove financially sound.
 - ☐ F. Place a check here if you think Mr. A should NOT take the new job no matter what the probabilities.
2. Mr. B, a 45 year old accountant, has recently been informed by his physician that he has developed a severe heart ailment. The disease would be sufficiently serious to force Mr. B to change many of his strongest life habits--reducing his work load, drastically changing his diet, giving up his favorite leisure-time pursuits. The physician suggests that a delicate medical operation could be attempted which, if successful, would completely relieve the heart condition. But, its success could

not be assured, and in fact, the operation might prove fatal.

Imagine that you are advising Mr. B. Listed below are several probabilities or odds of the operation will prove successful.

Please check the lowest probability that you would consider acceptable for the operation to be performed.

- ☐ A. Place a check here if you think Mr. B should NOT have the operation, no matter what the probabilities.
- ☐ B. The chances are 9 in 10 that the operation will be a success.
- ☐ C. The chances are 7 in 10 that the operation will be a success.
- ☐ D. The chances are 5 in 10 that the operation will be a success.
- ☐ E. The chances are 3 in 10 that the operation will be a success.
- ☐ F. The chances are 1 in 10 that the operation will be a success.

3. Mr. D is the captain of College X's football team. College X is playing traditional rival, College Y, in the final game of the season. The game few of the luxuries. Mr. C's father, who died recently, carried a \$4000 life insurance policy. Mr. C would like to invest this money in stocks. He is well aware of the secure "blue-chip" stocks and bonds that would pay approximately 6% on his investment. On the other hand, Mr. X might double their present value if a new product currently in production is favorably received by the buying public. However, if the product is unfavorably received, the stocks would decline in value.

Imagine that you are advising Mr. C. Listed below are several probabilities or odds that Company X stocks will double their value.

Please check the lowest probability that you would consider acceptable for Mr. C to invest in Company X Stocks.

- ☐ A. The chances are 1 in 10 that the stocks will double in their value.

- ☐ B. The chances are 3 in 10 that the stocks will double in their value.
- ☐ C. The chances are 5 in 10 that the stocks will double in their value.
- ☐ D. The chances are 7 in 10 that the stocks will double in their value.
- ☐ E. The chances are 9 in 10 that the stocks will double in their value.
- ☐ F. Place a check here if you think Mr. C should NOT invest in Company X stocks, no matter what the probabilities.

4. Mr. D is the captain of College X's football team. College X is playing traditional rival, College Y, in the final game of the season. The game is in its final seconds, and Mr. D's team, College X, is behind in the score. College X has time to run one more play. Mr. D, the captain, must decide whether it would be best to settle for a tie score with a play which would be almost certain to work or, on the other hand, should he try a more complicated and risky play which could bring victory if it succeeded, but defeat if not.

Imagine that you are advising Mr. D. Listed below are several probabilities or odds that the risky play will work.

Please check the lowest probability that you would consider acceptable for the risky play to be attempted.

- ☐ A. Place a check here if you think Mr. D should NOT attempt the risky play, no matter what the probabilities.
- ☐ B. The chances are 9 in 10 that the risky play will work.
- ☐ C. The chances are 7 in 10 that the risky play will work.
- ☐ D. The chances are 5 in 10 that the risky play will work.
- ☐ E. The chances are 3 in 10 that the risky play will work.
- ☐ E. The chances are 1 in 10 that the risky play will work.

5. Mr. E is the president of a light metals corporation in the United States. The corporation is quite prosperous, and has strongly considered the possibilities of business expansion by building an

additional plant in a new location. The choice is between building another plant in the U.S., where there would be a moderate return on the initial investment, or building a plant in a foreign country. Lower labor costs and easy access to raw materials in that country would mean a much higher return on the initial investment. On the other hand, there is a history of political instability and revolution in the foreign country under consideration. In fact, the leader of a small minority party is committed to nationalizing, that is, taking over, all foreign investments.

Imagine that you are advising Mr. E. Listed below are several probabilities or odds of continued political stability in the foreign country under consideration.

Please check the lowest probability that you would consider acceptable for Mr. E's corporation to build a plant in that country.

- ☐ A. The chances are 1 in 10 that the foreign country will remain politically stable.
 - ☐ B. The chances are 3 in 10 that the foreign country will remain politically stable.
 - ☐ C. The chances are 5 in 10 that the foreign country will remain politically stable.
 - ☐ D. The chances are 7 in 10 that the foreign country will remain politically stable.
 - ☐ E. The chances are 9 in 10 that the foreign country will remain politically stable.
 - ☐ F. Place a check here if you think Mr. E's corporation should NOT build a plant in the foreign country, no matter what the probabilities.
6. Mr. F is currently a college senior who is very eager to pursue graduate study in chemistry, leading to the Doctor of Philosophy degree. He has been accepted by both University X and University Y. University X has a world-wide reputation for excellence in chemistry. While a degree from University X would signify outstanding training in this field, the standards are so very rigorous that only a fraction of the degree candidates actually receive the degree. University Y, on the other hand, has much less of a reputation in chemistry, but almost everyone admitted is awarded the Doctor of Philosophy degree though the degree has much less

prestige than the corresponding degree from University X.

Imagine that you are advising Mr. F. Listed below are several probabilities or odds that Mr. F would be awarded a degree at University X, the one with the greater prestige.

Please check the lowest probability that you would consider acceptable to make it worthwhile for Mr. F to enroll in University X rather than University Y.

- ☐ A. Place a check here if you think Mr. F should NOT enroll in University X, no matter what the probabilities.
- ☐ B. The chances are 9 in 10 that Mr. F would receive a degree from University X.
- ☐ C. The chances are 7 in 10 that Mr. F would receive a degree from University X.
- ☐ D. The chances are 5 in 10 that Mr. F would receive a degree from University X.
- ☐ E. The chances are 3 in 10 that Mr. F would receive a degree from University X.
- ☐ F. The chances are 1 in 10 that Mr. F would receive a degree from University X.

7. Mr. G, a competent chess player, is participating in a national chess tournament. In an early match he draws the top-favored player in the tournament as his opponent. Mr. G has been given a relatively low ranking in view of his performance in previous tournaments. During the course of his play with the top-favored man, Mr. G notes the possibility of a deceptive though risky maneuver which might bring him a quick victory. At the same time, if the attempted maneuver should fail, Mr. G would be left in an exposed position and defeat would almost certainly follow.

Imagine that you are advising Mr. G. Listed below are several probabilities or odds that Mr. G's deceptive play would succeed.

Please check the lowest probability that you would consider acceptable for the risky play in question to be attempted.

- ☐ A. The chances are 1 in 10 that the play would succeed.

- ___ B. The chances are 3 in 10 that the play would succeed.
- ___ C. The chances are 5 in 10 that the play would succeed.
- ___ D. The chances are 7 in 10 that the play would succeed.
- ___ E. The chances are 9 in 10 that the play would succeed.
- ___ F. Place a check here if you think Mr. G should NOT attempt the risky play, no matter what the probabilities.

8. Mr. H, a college senior, has studied the piano since childhood. He has won amateur prizes and given small recitals, suggesting that Mr. H has considerable musical talent. As graduation approaches, Mr. H has the choice of going to medical school to become a physician, a profession which would bring certain prestige and financial rewards; or entering a conservatory of music for advanced training with a well-known pianist. Mr. H realizes that even upon completion of his piano studies, which would take many more years and a lot of money, success as a concert pianist would not be assured.

Imagine that you are advising Mr. H. Listed below are several probabilities or odds that Mr. H would succeed as a concert pianist.

Please check the lowest probability that you would consider acceptable for Mr. H to continue with his musical training.

- ___ A. Place a check here if you think Mr. H should NOT pursue his musical training, no matter what the probabilities.
- ___ B. The chances are 9 in 10 that Mr. H would succeed as a concert pianist.
- ___ C. The chances are 9 in 10 that Mr. H would succeed as a concert pianist.
- ___ D. The chances are 9 in 10 that Mr. H would succeed as a concert pianist.
- ___ E. The chances are 9 in 10 that Mr. H would succeed as a concert pianist.
- ___ F. The chances are 9 in 10 that Mr. H would succeed as a concert pianist.

9. Mr. J is an American captured by the enemy in World War II and placed in a prisoner-of-war camp. Conditions in the camp are quite bad, with long

hours of hard physical labor and a barely sufficient diet. After spending several months in this camp, Mr. J notes the possibility of escape by concealing himself in a supply truck that shuttles in and out of the camp. Of course, there is no guarantee that the escape would prove successful. Recapture by the enemy could well mean execution.

Imagine that you are advising Mr. J. Listed below are several probabilities or odds of a successful escape from the prisoner-of-war camp.

Please check the lowest probability that you would consider acceptable for an escape to be attempted.

- ☐ A. The chances are 1 in 10 that the escape would succeed.
- ☐ B. The chances are 3 in 10 that the escape would succeed.
- ☐ C. The chances are 5 in 10 that the escape would succeed.
- ☐ D. The chances are 7 in 10 that the escape would succeed.
- ☐ E. The chances are 9 in 10 that the escape would succeed.
- ☐ F. Place a check here if you think Mr. H should NOT try to escape, no matter what the probabilities.

10. Mr. K is a successful businessman who has participated in a number of civic activities of considerable value to the community. Mr. K has been approached by the leaders of his political party as a possible congressional candidate in the next election. Mr. K's party is a minority party in the district, though the party has won occasional elections in the past. Mr. K would like to hold political office, but to do so would involve a serious financial sacrifice, since the party has insufficient campaign funds. He would also have to endure the attacks of his political opponents in a hot campaign.

Imagine that you are advising Mr. K. Listed below are several probabilities or odds of Mr. K's winning the election in his district.

Please check the lowest probability that you would consider acceptable to make it worthwhile for Mr. K to run for political office.

- ☐ A. Place a check here if you think Mr. K should NOT run for political office, no matter what the probabilities.
- ☐ B. The chances are 9 in 10 that Mr. K would win the election.
- ☐ C. The chances are 7 in 10 that Mr. K would win the election.
- ☐ D. The chances are 5 in 10 that Mr. K would win the election.
- ☐ E. The chances are 3 in 10 that Mr. K would win the election.
- ☐ F. The chances are 1 in 10 that Mr. K would win the election.

11. Mr. L, a married 30 year-old research physicist, has been given a five-year appointment by a major university laboratory. As he contemplates the next five years, he realizes that he might work on a difficult, long-term problem which, if a solution could be found, would resolve basic scientific issues in the field and bring high scientific honors. If no solution were found, however, Mr. L would have little to show for his five years in the laboratory, and this would make it hard for him to get a good job afterwards. On the other hand, he could, as most of his professional associates are doing, work on a series of short-term problems where solutions would be easier to find, but where the problems are of lesser scientific importance.

Imagine that you are advising Mr. L. Listed below are several probabilities or odds that a solution would be found to the difficult, long-term problem that Mr. L has in mind.

Please check the lowest probability that you would consider acceptable to make it worthwhile for Mr. L to work on the more difficult long-term problem.

- ☐ A. The chances are 1 in 10 that Mr. L would solve the long-term problem.
- ☐ B. The chances are 3 in 10 that Mr. L would solve the long-term problem.
- ☐ C. The chances are 5 in 10 that Mr. L would solve the long-term problem.
- ☐ D. The chances are 7 in 10 that Mr. L would solve the long-term problem.
- ☐ E. The chances are 9 in 10 that Mr. L would solve the long-term problem.

- F. Place a check here if you think Mr. L should NOT choose the long-term, difficult problem, no matter what the probabilities.

12. Mr. M is contemplating marriage to Miss T, a woman whom he has known a little more than a year. Recently, however, a number of arguments have occurred between them, suggesting some sharp differences of opinion in the way each views certain matters. Indeed, they decide to seek professional advice from a marriage counselor as to whether it would be wise for them to marry. On the basis of these meetings with a marriage counselor, they realize that a happy marriage, while possible, would not be assured.

Imagine that you are advising Mr. M and Miss T. Listed below are several probabilities or odds that their marriage would prove to be a happy and successful one.

Please check the lowest probability that you would consider acceptable for Mr. M and Miss T. to get married.

- A. Place a check here if you think Mr. M and Miss T should NOT marry, no matter what the probabilities.
- B. The chances are 9 in 10 that the marriage would be happy and successful.
- C. The chances are 7 in 10 that the marriage would be happy and successful.
- D. The chances are 5 in 10 that the marriage would be happy and successful.
- E. The chances are 3 in 10 that the marriage would be happy and successful.
- F. The chances are 1 in 10 that the marriage would be happy and successful.

APPENDIX F

Verbal Instructions Given For Sample SAT Test

50 ITEM MULTIPLE CHOICE TEST INSTRUCTIONS (SAT)

On the following pages, you will find a series of questions. There are three types: analogy questions, mathematical questions, and antonym questions.

For the analogy questions, a related pair of words is followed by five lettered pairs of words. Select the lettered pair that best expresses a relationship similar to that expressed in the original pair.

Antonym questions consist of a word printed in capital letters, followed by five lettered words. Choose the lettered word that is most nearly opposite in meaning to the word in capital letters.

For those mathematical questions, select the best one of the five choices available.

There are 50 questions in all. Each question has only one correct answer. Please answer all questions. Are there any questions concerning these instructions? Please begin. You have 45 minutes to complete this test.

APPENDIX G

Verbal Instructions Given For Risk-Taking Test

RISK-TAKING TEST INSTRUCTIONS

On the following pages, you will find a series of situations that are likely to occur in everyday life. The central person in each situation is faced with a choice between two alternative courses of action, which we might call X and Y. Alternative X is more desirable and attractive than Alternative Y, but the probability of attaining or achieving X is less than that of attaining or achieving Y.

For each situation on the following pages, you will be asked to indicate the minimum odds of success you would demand before recommending that the more attractive or desirable alternative X, be chosen.

Read each situation carefully before giving your judgment. Try to place yourself in the position of the central person in each of the situations. There are twelve situations in all. Please do not omit any of them.

NOTE: This Opinion Questionnaire II (Choice Dilemmas Procedure/Risk-Taking Test) was extracted from Appendix E of "Risk Taking: A Study in Cognition and Personality," written by N. Kogan, and M. Wallach, 1964, Holt, Rinehart and Winston.

APPENDIX H
Self-Assessment Instructions

SELF-ASSESSMENT INSTRUCTIONS

On this test you first select an answer and then indicate HOW SURE YOU ARE that your answer is correct.

Your test score depends on:

1. The CORRECTNESS of your answer, and
2. You can obtain bonus points for the ACCURACY of your confidence assessment.

Read each question carefully, try to answer them as correctly as you can, and self-assess immediately after each question.

It is important to note that the self-assessment scale asks you HOW SURE you are that your answer to the question is "correct."

You get POINTS for giving a CORRECT ANSWER.

You get BONUS POINTS for making an ACCURATE CONFIDENCE ASSESSMENT!

So the more accurate your confidence assessments, the higher your score on the test.

The particular points for scoring have been selected so that YOU WILL OBTAIN THE HIGHEST SCORE BY ACCURATELY AND TRUTHFULLY INDICATING "HOW SURE" YOU ARE.

APPENDIX I
Self-Assessment Answer Sheet
(1990 Version)

MULTIPLE CHOICE SELF ASSESSMENT ANSWER SHEET

HUMAN PERFORMANCE ENHANCEMENT, INC.
SELF ASSESSMENT TECHNOLOGIES
© COPYRIGHT 1990
ALL RIGHTS RESERVED

DIRECTIONS:

1. USE A NO. 2 PENCIL ONLY.
2. ONLY ONE ANSWER PER QUESTION ALLOWED.
3. MAKE NO STRAY MARKS ON THIS SHEET.
4. ERASE CLEAN ANY MARK YOU WISH TO CHANGE.
5. DO NOT FOLD OR STAPLE THIS SHEET.

EXAMPLE: T F ANSWER HOW SURE ARE YOU?
A B C ☒ D E F G H I J 0 1 2 3 4
YOUR ANSWER IS **D** AND YOU ARE VERY SURE **3**
THAT YOUR ANSWER IS CORRECT

WRITE SOCIAL SECURITY NUMBER IN
SPACE PROVIDED BLACKEN IN CIRCLE BELOW
CORRESPONDING TO NUMBER ENTERED

0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

DO NOT MARK IN THIS SPACE — FOR COMPUTER
CENTER USE ONLY

NAME _____ DATE _____

COURSE TITLE _____ INSTRUCTOR _____

		HOW SURE ARE YOU?							HOW SURE ARE YOU?							
		NOT SURE AT ALL	VERY UNSURE	SOMEWHAT SURE	VERY SURE	EXTREMELY SURE			NOT SURE AT ALL	VERY UNSURE	SOMEWHAT SURE	VERY SURE	EXTREMELY SURE			
ANSWER							ANSWER									
T F	1	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	2	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	3	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	4	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	5	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	6	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	7	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	8	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	9	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	10	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	11	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	12	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	13	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	14	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	15	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	16	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	17	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	18	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	19	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	20	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	21	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	22	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	23	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	24	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	25	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	26	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	27	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	28	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	29	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4
T F	30	A	B	C	D	E	F	G	H	I	J	0	1	2	3	4

**DO NOT MARK
OR WRITE
IN THIS SPACE**

		HOW SURE ARE YOU?							HOW SURE ARE YOU?				
ANSWER		NOT SURE AT ALL	VERY UNSURE	SOMEWHAT SURE	VERY SURE	EXTREMELY SURE	ANSWER		NOT SURE AT ALL	VERY UNSURE	SOMEWHAT SURE	VERY SURE	EXTREMELY SURE
31	T F						48	T F					
32	T F						49	T F					
33	T F						50	T F					
34	T F						51	T F					
35	T F						52	T F					
36	T F						53	T F					
37	T F						54	T F					
38	T F						55	T F					
39	T F						56	T F					
40	T F						57	T F					
41	T F						58	T F					
42	T F						59	T F					
43	T F						60	T F					
44	T F						61	T F					
45	T F						62	T F					
46	T F						63	T F					
47	T F						64	T F					
48	T F						65	T F					
49	T F						66	T F					
50	T F						67	T F					
51	T F						68	T F					
52	T F						69	T F					
53	T F						70	T F					
54	T F						71	T F					
55	T F						72	T F					
56	T F						73	T F					
57	T F						74	T F					
58	T F						75	T F					
59	T F						76	T F					
60	T F						77	T F					
61	T F						78	T F					
62	T F						79	T F					
63	T F						80	T F					
64	T F						81	T F					
65	T F						82	T F					
66	T F						83	T F					
67	T F						84	T F					
68	T F						85	T F					
69	T F						86	T F					
70	T F						87	T F					
71	T F						88	T F					
72	T F						89	T F					
73	T F						90	T F					
74	T F						91	T F					
75	T F						92	T F					
76	T F						93	T F					
77	T F						94	T F					
78	T F						95	T F					
79	T F						96	T F					
80	T F						97	T F					
81	T F						98	T F					
82	T F						99	T F					
83	T F						100	T F					
84	T F												
85	T F												
86	T F												
87	T F												
88	T F												
89	T F												
90	T F												
91	T F												
92	T F												
93	T F												
94	T F												
95	T F												
96	T F												
97	T F												
98	T F												
99	T F												
100	T F												

APPENDIX J

Usability Answer Sheet

MULTIPLE CHOICE SELF ASSESSMENT ANSWER SHEET

HUMAN PERFORMANCE ENHANCEMENT INC.
SELF ASSESSMENT TECHNOLOGIES
© COPYRIGHT 1990
ALL RIGHTS RESERVED

WRITE SOCIAL SECURITY NUMBER IN
SPACE PROVIDED. BLACKEN IN CIRCLE BELOW
CORRESPONDING TO NUMBER ENTERED.

DIRECTIONS:

1. USE A NO. 2 PENCIL ONLY
2. ONLY ONE ANSWER PER QUESTION ALLOWED
3. MAKE NO STRAY MARKS ON THIS SHEET
4. ERASE CLEAN ANY MARK YOU WISH TO CHANGE
5. DO NOT FOLD OR STAPLE THIS SHEET

0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9

DO NOT MARK IN THIS SPACE — FOR COMPUTER
CENTER USE ONLY

NAME _____ DATE _____

COURSE TITLE _____ INSTRUCTOR _____

		HOW USEFUL IS THIS?							HOW USEFUL IS THIS?				
		NOT USEFUL AT ALL	NOT VERY USEFUL	SOMEWHAT USEFUL	VERY USEFUL	EXTREMELY USEFUL			NOT USEFUL AT ALL	NOT VERY USEFUL	SOMEWHAT USEFUL	VERY USEFUL	EXTREMELY USEFUL
ANSWER							ANSWER						
1	T F	A	B	C	D	E	0	1	2	3	4		
2	T F	A	B	C	D	E	0	1	2	3	4		
3	T F	A	B	C	D	E	0	1	2	3	4		
4	T F	A	B	C	D	E	0	1	2	3	4		
5	T F	A	B	C	D	E	0	1	2	3	4		
6	T F	A	B	C	D	E	0	1	2	3	4		
7	T F	A	B	C	D	E	0	1	2	3	4		
8	T F	A	B	C	D	E	0	1	2	3	4		
9	T F	A	B	C	D	E	0	1	2	3	4		
10	T F	A	B	C	D	E	0	1	2	3	4		
11	T F	A	B	C	D	E	0	1	2	3	4		
12	T F	A	B	C	D	E	0	1	2	3	4		
13	T F	A	B	C	D	E	0	1	2	3	4		
14	T F	A	B	C	D	E	0	1	2	3	4		
15	T F	A	B	C	D	E	0	1	2	3	4		
16	T F	A	B	C	D	E	0	1	2	3	4		
17	T F	A	B	C	D	E	0	1	2	3	4		
18	T F	A	B	C	D	E	0	1	2	3	4		
19	T F	A	B	C	D	E	0	1	2	3	4		
20	T F	A	B	C	D	E	0	1	2	3	4		
21	T F	A	B	C	D	E	0	1	2	3	4		
22	T F	A	B	C	D	E	0	1	2	3	4		
23	T F	A	B	C	D	E	0	1	2	3	4		
24	T F	A	B	C	D	E	0	1	2	3	4		
25	T F	A	B	C	D	E	0	1	2	3	4		
26	T F	A	B	C	D	E	0	1	2	3	4		
27	T F	A	B	C	D	E	0	1	2	3	4		
28	T F	A	B	C	D	E	0	1	2	3	4		
29	T F	A	B	C	D	E	0	1	2	3	4		
30	T F	A	B	C	D	E	0	1	2	3	4		

DO NOT MARK
OR WRITE
IN THIS SPACE

		HOW USEFUL IS THIS?							HOW USEFUL IS THIS?				
		NOT USEFUL AT ALL	NOT VERY USEFUL	SOMEWHAT USEFUL	VERY USEFUL	EXTREMELY USEFUL			NOT USEFUL AT ALL	NOT VERY USEFUL	SOMEWHAT USEFUL	VERY USEFUL	EXTREMELY USEFUL
ANSWER							ANSWER						
31	T F A B C D E F G H I J	0	1	2	3	4	46	T F A B C D E F G H I J	0	1	2	3	4
32	T F A B C D E F G H I J	0	1	2	3	4	47	T F A B C D E F G H I J	0	1	2	3	4
33	T F A B C D E F G H I J	0	1	2	3	4	48	T F A B C D E F G H I J	0	1	2	3	4
34	T F A B C D E F G H I J	0	1	2	3	4	49	T F A B C D E F G H I J	0	1	2	3	4
35	T F A B C D E F G H I J	0	1	2	3	4	50	T F A B C D E F G H I J	0	1	2	3	4
36	T F A B C D E F G H I J	0	1	2	3	4	51	T F A B C D E F G H I J	0	1	2	3	4
37	T F A B C D E F G H I J	0	1	2	3	4	52	T F A B C D E F G H I J	0	1	2	3	4
38	T F A B C D E F G H I J	0	1	2	3	4	53	T F A B C D E F G H I J	0	1	2	3	4
39	T F A B C D E F G H I J	0	1	2	3	4	54	T F A B C D E F G H I J	0	1	2	3	4
40	T F A B C D E F G H I J	0	1	2	3	4	55	T F A B C D E F G H I J	0	1	2	3	4
41	T F A B C D E F G H I J	0	1	2	3	4	56	T F A B C D E F G H I J	0	1	2	3	4
42	T F A B C D E F G H I J	0	1	2	3	4	57	T F A B C D E F G H I J	0	1	2	3	4
43	T F A B C D E F G H I J	0	1	2	3	4	58	T F A B C D E F G H I J	0	1	2	3	4
44	T F A B C D E F G H I J	0	1	2	3	4	59	T F A B C D E F G H I J	0	1	2	3	4
45	T F A B C D E F G H I J	0	1	2	3	4	60	T F A B C D E F G H I J	0	1	2	3	4

APPENDIX K

Instructions For Usability Answer Sheet

ASSESSMENT INSTRUCTIONS

On this test you first, select an answer and then indicate how useful you think this information (the actual question) is for you to know as a college freshman.

Your test score depends on:

1. The CORRECTNESS of your answer, and
2. You can obtain bonus points for the ACCURACY of your "USEFULNESS" assessment.

Read each question carefully, try to answer them as correctly as you can, and self-assess immediately after each question.

You get POINTS for giving a CORRECT ANSWER.

You get BONUS POINTS for making an ACCURATE "USEFULNESS" ASSESSMENT!

So the more accurate your confidence assessments, the higher your score on the test.

The particular points for scoring have been selected so that YOU WILL OBTAIN THE HIGHEST SCORE BY ACCURATELY AND TRUTHFULLY INDICATING HOW USEFUL YOU THINK THE INFORMATION IS.

APPENDIX L

Instructions for Multiple-Choice Answer Sheet

MULTIPLE-CHOICE ANSWER SHEET INSTRUCTIONS

Please read each question carefully and then mark your answer on the blue answer sheet provided.

1. Only one response per question allowed.
2. Make no stray marks on this sheet.
3. Erase clean any mark you wish to change.
4. Do not fold or staple this sheet.
5. REMEMBER, THERE ARE 50 QUESTIONS ON THIS TEST!

APPENDIX M

Supplementary Figures and Tables

Appendix Table M1

**Means and Variances of Number of
Correct Responses for Treatment, Ethnicity, and
Gender Based on 20 Observations Per Cell**

TRT ¹	ETH ²	GENDER	MEAN	VARIANCE
UA	NH	M	29.3	41.6
		F	30.2	44.5
	H	M	23.2	31.8
		F	18.6	35.9
SA	NH	M	26.4	61.9
		F	26.6	34.0
	H	M	21.0	39.8
		F	19.7	32.4
NOSA	NH	M	27.3	34.5
		F	27.9	43.1
	H	M	26.2	68.9
		F	21.4	43.7
¹ Treatment UA = Usability Assessment SA = Self-Assessment NOSA = No Self-Assessment				
² Ethnicity NH = Non-Hispanic H = Hispanic				
Variances are Homogeneous. Bartlett's Test for Homogeneity of Variance resulted in a test statistic (χ^2) of 6.53, ($p = 0.83$)				

Appendix Table M2

ANCOVA Table Showing p Values
 Calculated for Number of Correct Responses
 For Gender, Ethnicity, and Treatment
 Based on 20 Observations Per Cell

SOURCE	Df	MS	F	p
GPA*	1	923.91	23.79	0.0001
Gender	1	246.95	6.36	0.0124
Ethnicity	1	1415.80	36.45	0.0001
Treatment	2	134.72	3.47	0.0328
Gender * Ethnicity	1	184.49	4.75	0.0303
Gender * Treatment	2	18.51	0.48	0.6215
Ethnicity * Treatment	2	138.82	3.57	0.0296
Gender * Ethnicity * Treatment	2	37.42	0.96	0.3831
Error	227			
TOTAL	239			
* Covariate				
MSE = 38.8				

Appendix Table M-3
Mean Risk Scores For Treatment,
Ethnicity, and Gender Based on 20 Observations Per Cell

TRT ¹	ETH ²	GENDER	MEAN
UA	NH	M	69.6
		F	67.5
	H	M	65.6
		F	70.3
SA	NH	M	68.3
		F	66.1
	H	M	68.8
		F	66.5
NOSA	NH	M	62.9 ^a
		F	68.4
	H	M	65.0
		F	78.1 ^b

¹Treatment

UA = Usability Assessment

SA = Self-Assessment

NOSA = No Self-Assessment

²Ethnicity

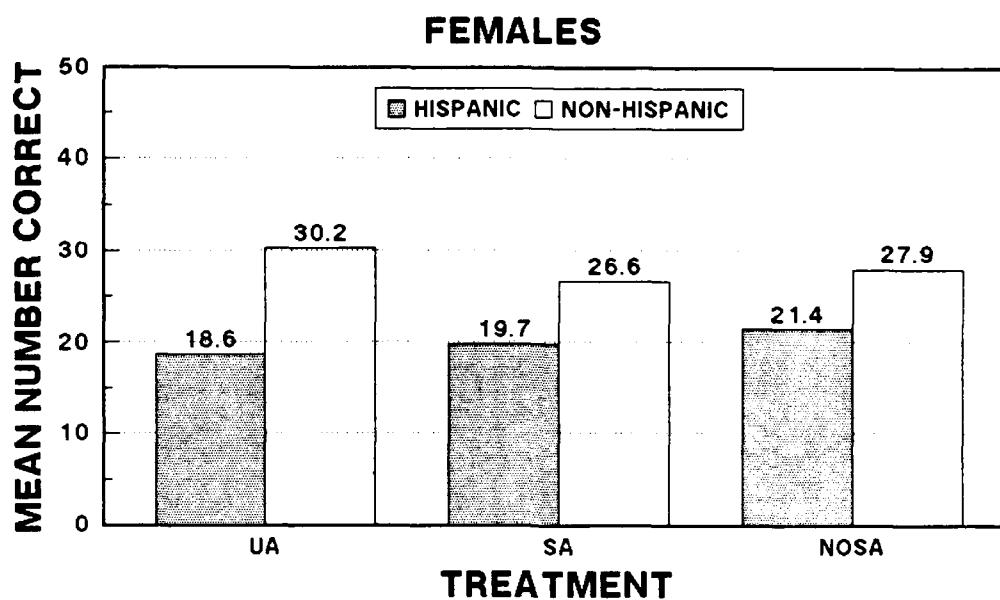
NH = Non-Hispanic

H = Hispanic

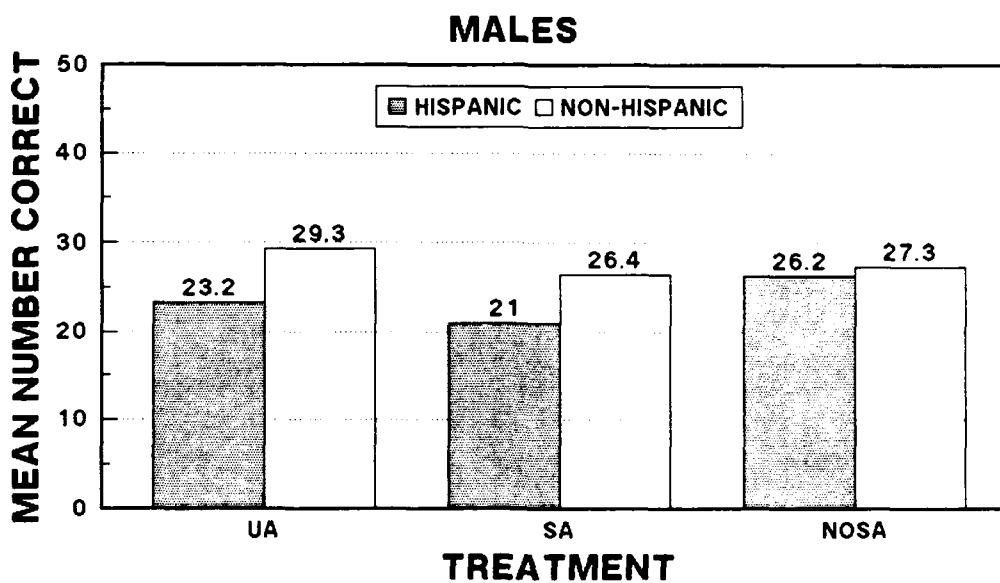
Kruskal-Wallis Procedure resulted in a test statistic (χ^2) of 19.8, ($p < .05$).

^aMean for Non-Hispanic males tested without self-assessing was significantly different from **b**, but not significantly different from rest of groups.

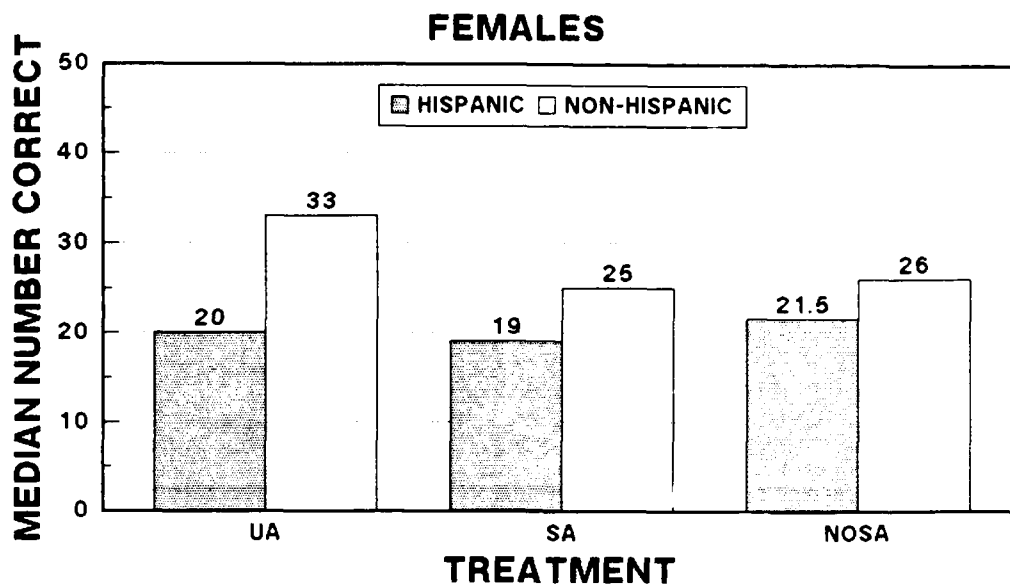
^bMean for Hispanic females tested without self-assessing was significantly different from **a**, but not significantly different from rest of groups.



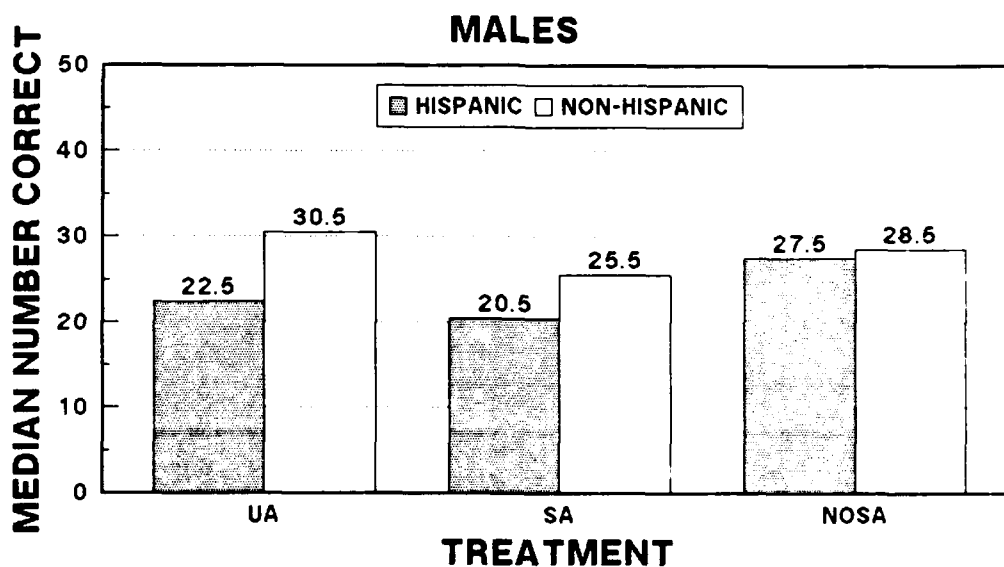
Appendix Figure M1. Mean number correct for each treatment for Non-Hispanic and Hispanic females.



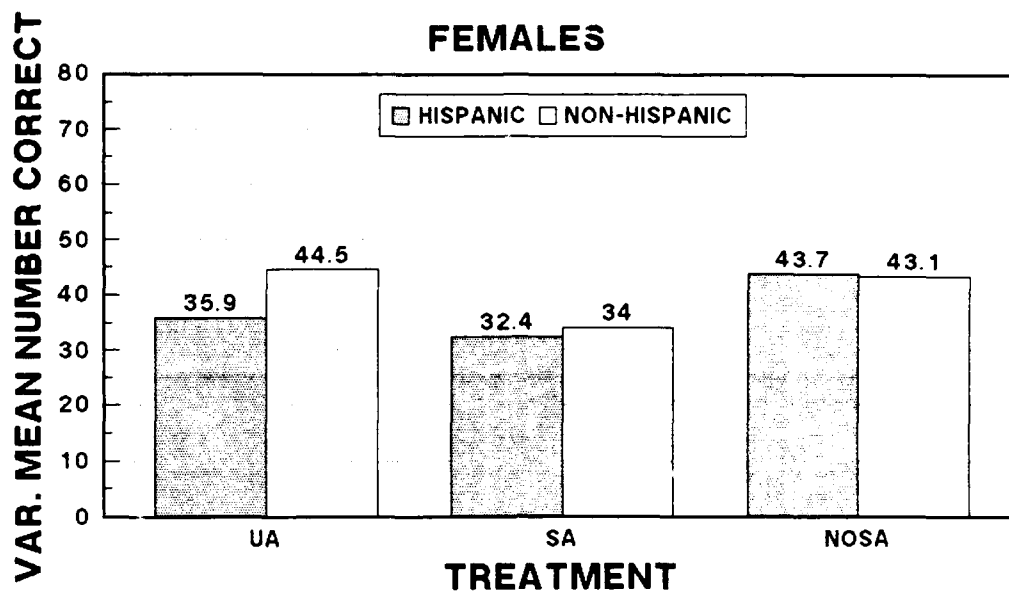
Appendix Figure M2. Mean number correct for each treatment for Non-Hispanic and Hispanic males.



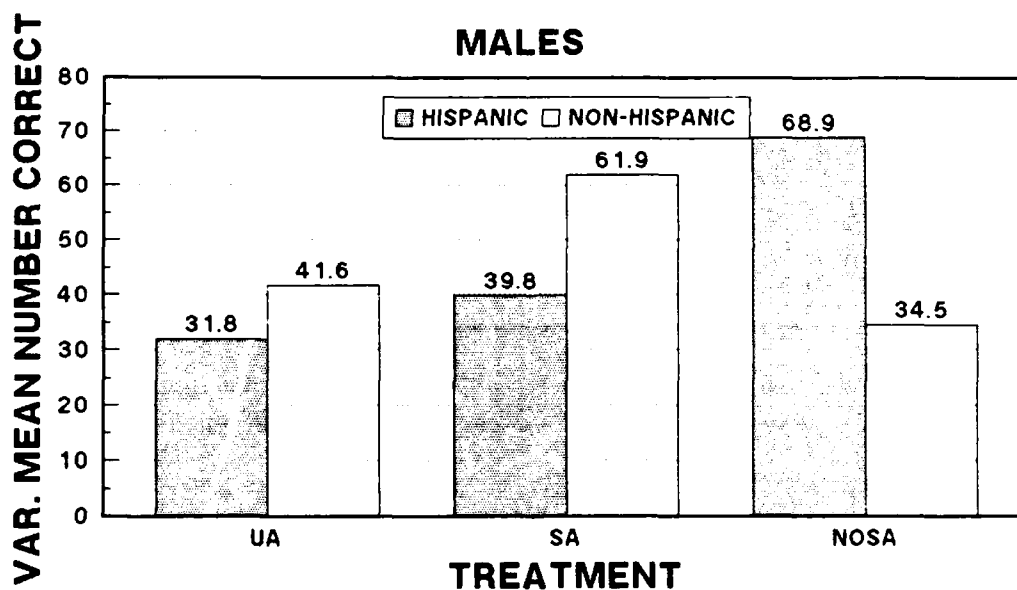
Appendix Figure M3. Median number correct for each treatment for Non-Hispanic and Hispanic females.



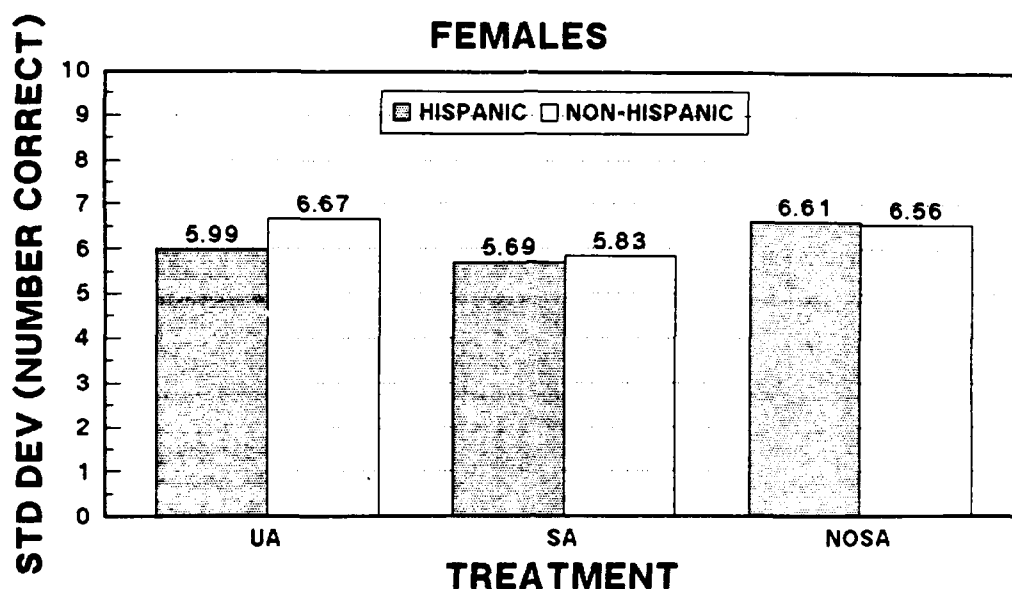
Appendix Figure M4. Median number correct for each treatment for Non-Hispanic and Hispanic males.



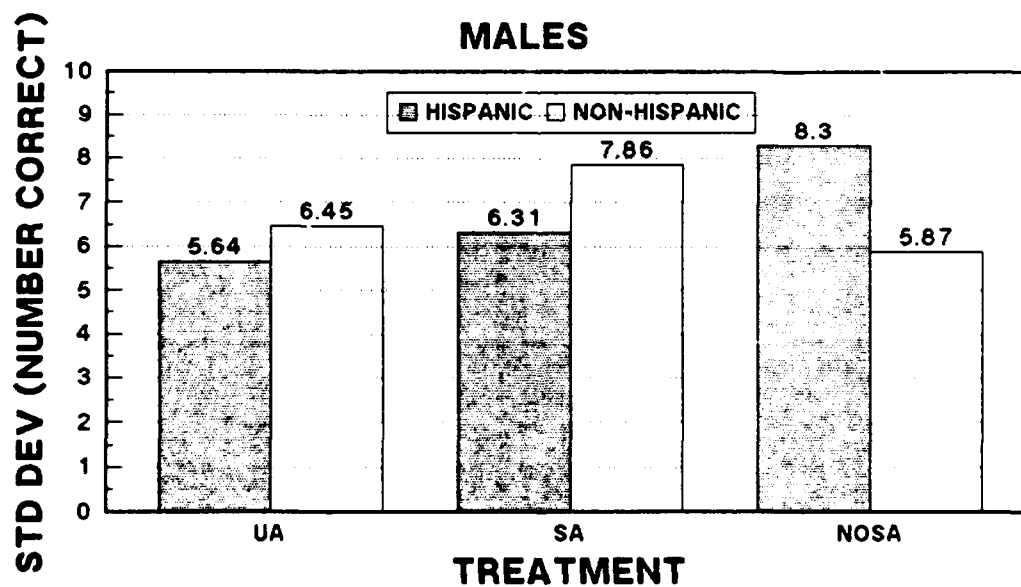
Appendix Figure M5. Variances for mean number correct for each treatment for Non-Hispanic and Hispanic females.



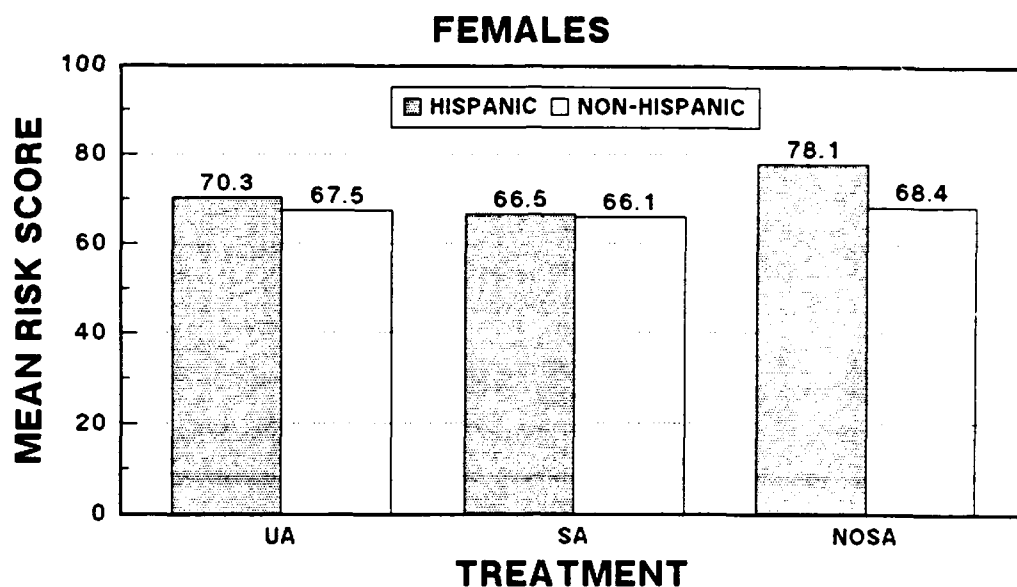
Appendix Figure M6. Variances for mean number correct for each treatment for Non-Hispanic and Hispanic males.



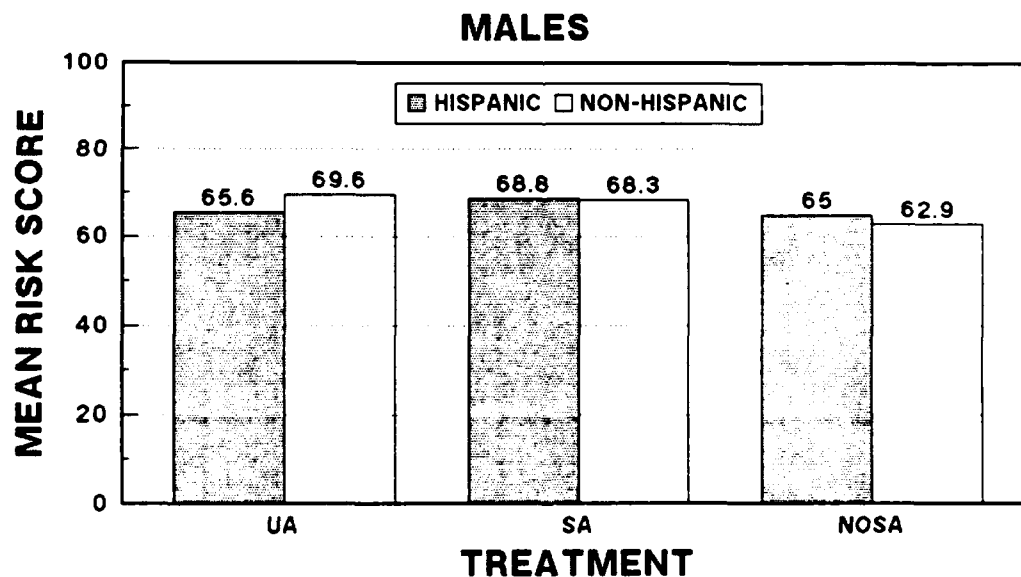
Appendix Figure M7. Standard deviations for mean number correct for each treatment for Non-Hispanic and Hispanic females.



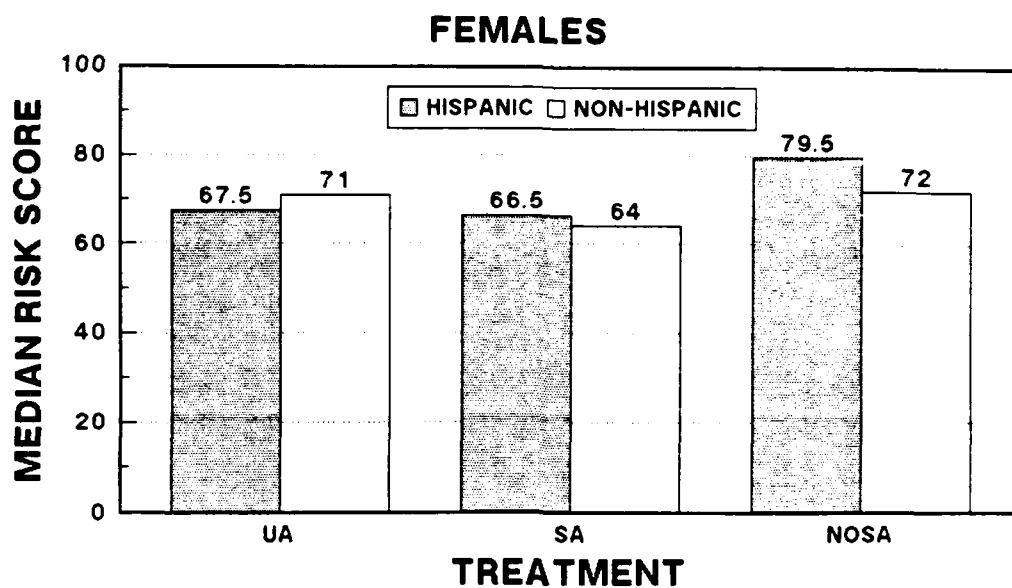
Appendix Figure M8. Standard deviations for mean number correct for each treatment for Non-Hispanic and Hispanic males.



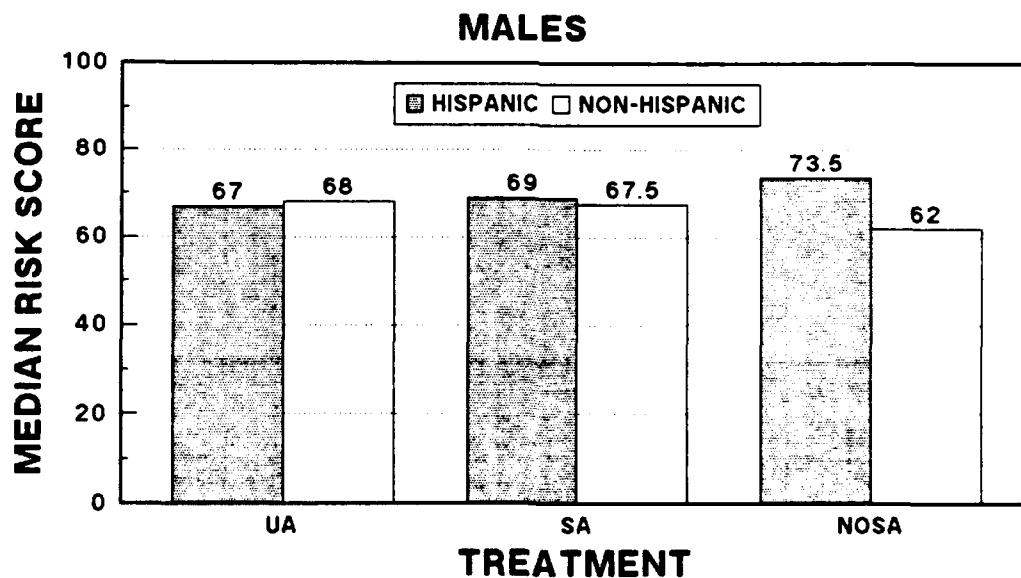
Appendix Figure M9. Mean risk score for each treatment for Non-Hispanic and Hispanic females.



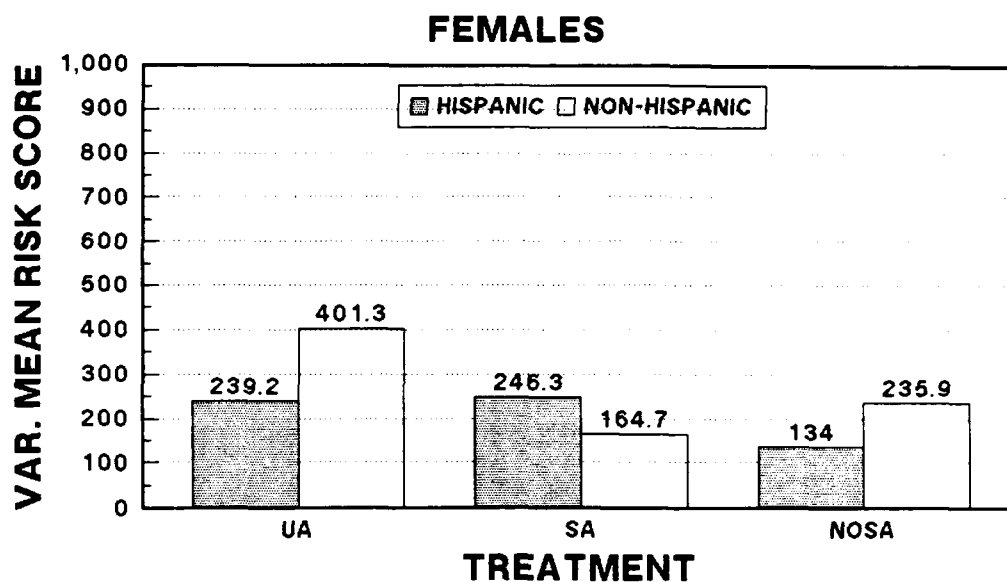
Appendix Figure M10. Mean risk score for each treatment for Non-Hispanic and Hispanic males.



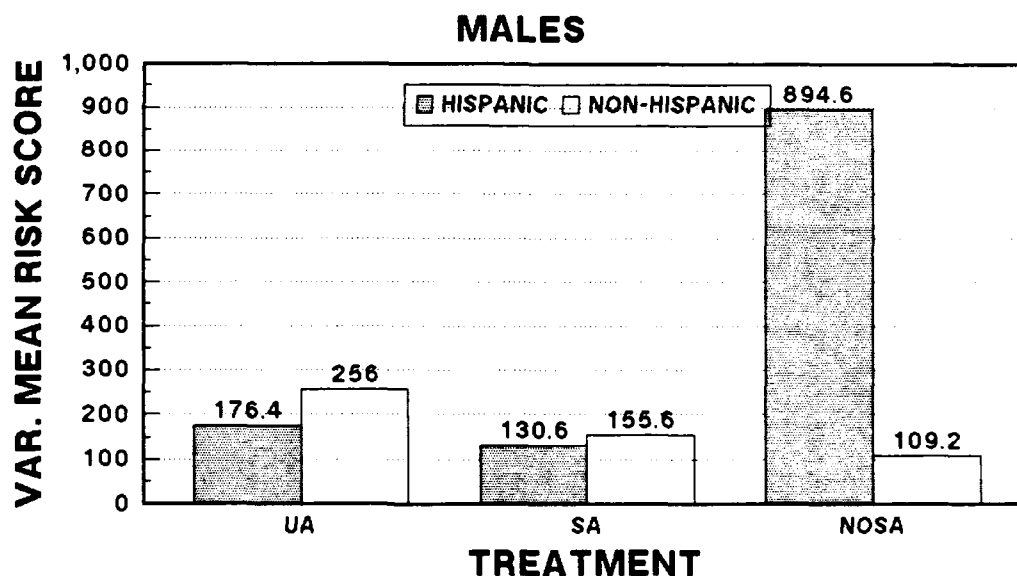
Appendix Figure M 11. Median risk score for each treatment for Non-Hispanic and Hispanic females.



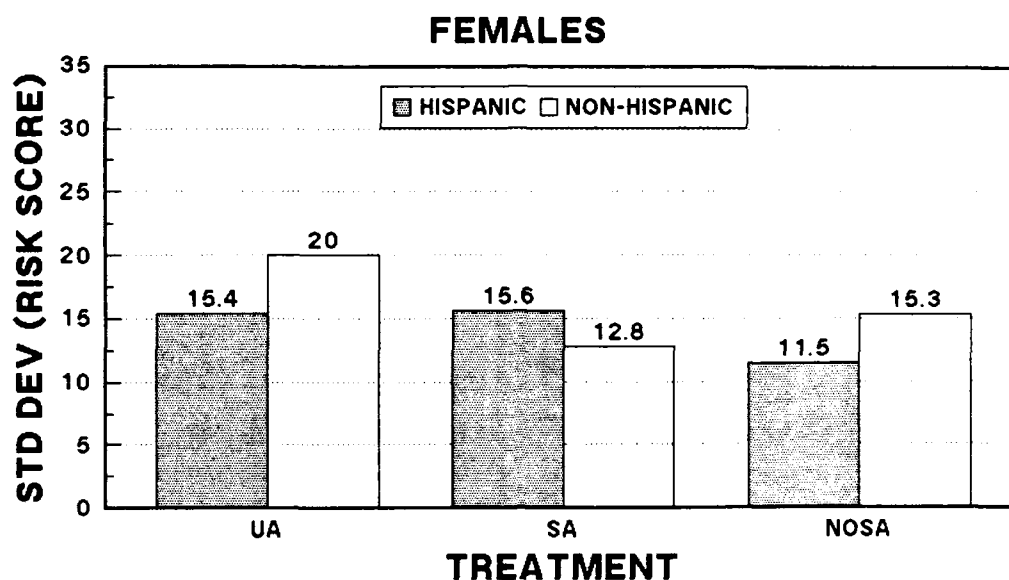
Appendix Figure M 12. Median risk score for each treatment for Non-Hispanic and Hispanic males.



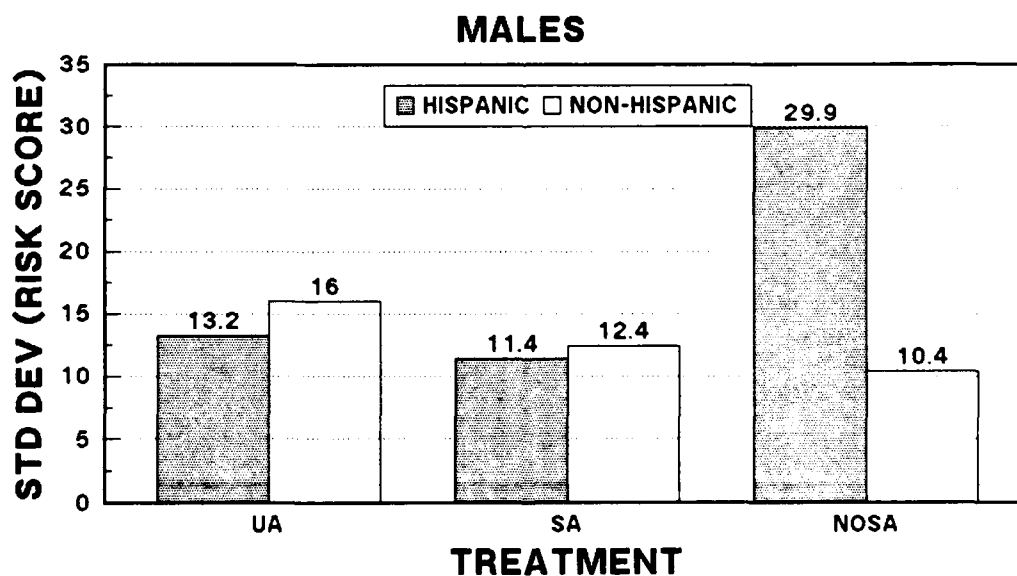
Appendix Figure M13. Variances for mean risk score for each treatment for Non-Hispanic and Hispanic females.



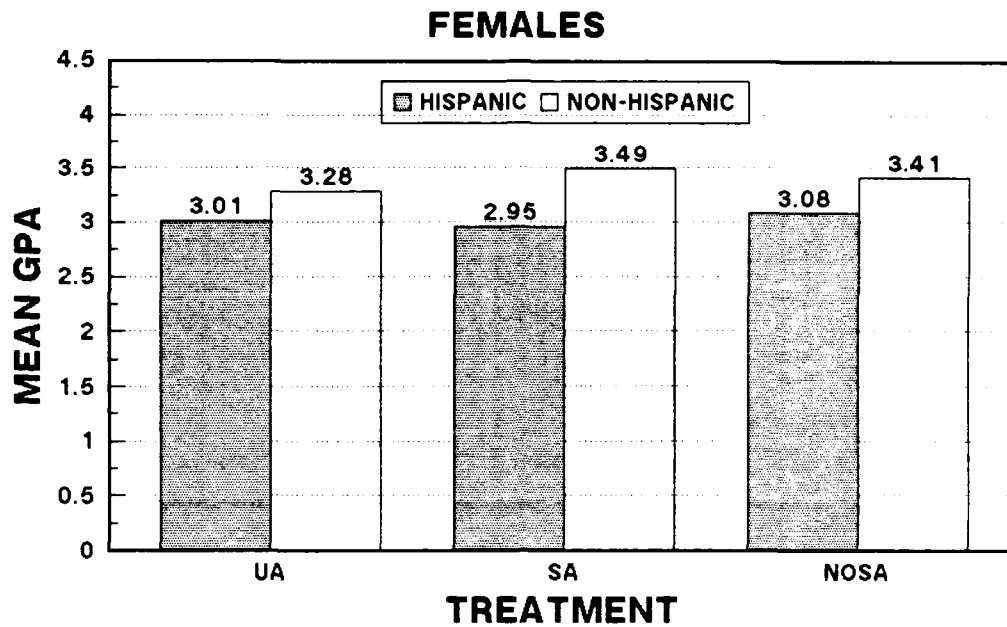
Appendix Figure M14. Variances for mean risk score for treatment for Non-Hispanic and Hispanic males.



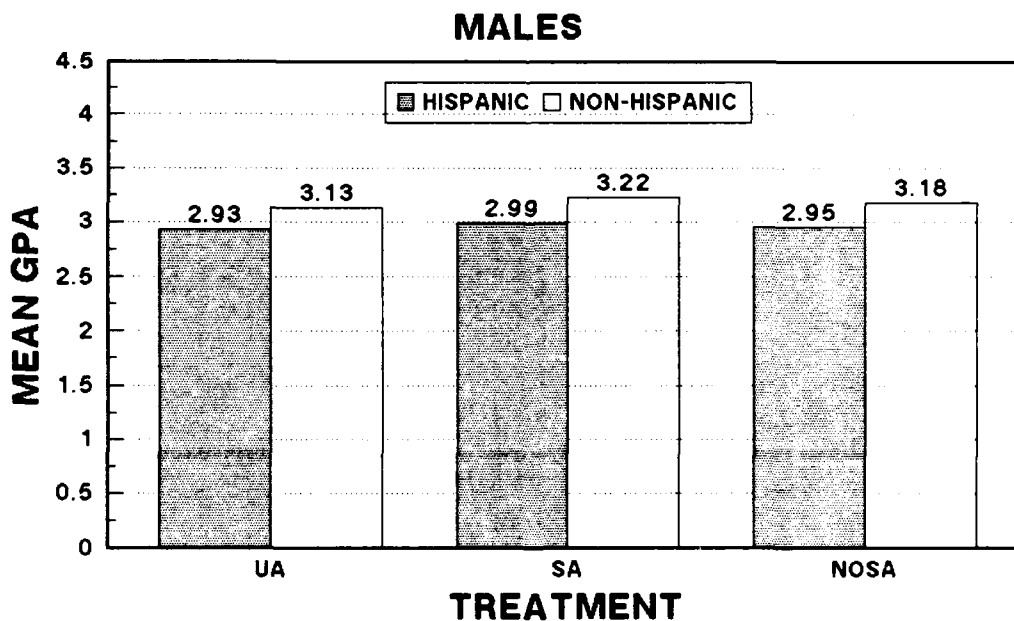
Appendix Figure M15. Standard deviations for mean risk score for each treatment for Non-Hispanic and Hispanic females.



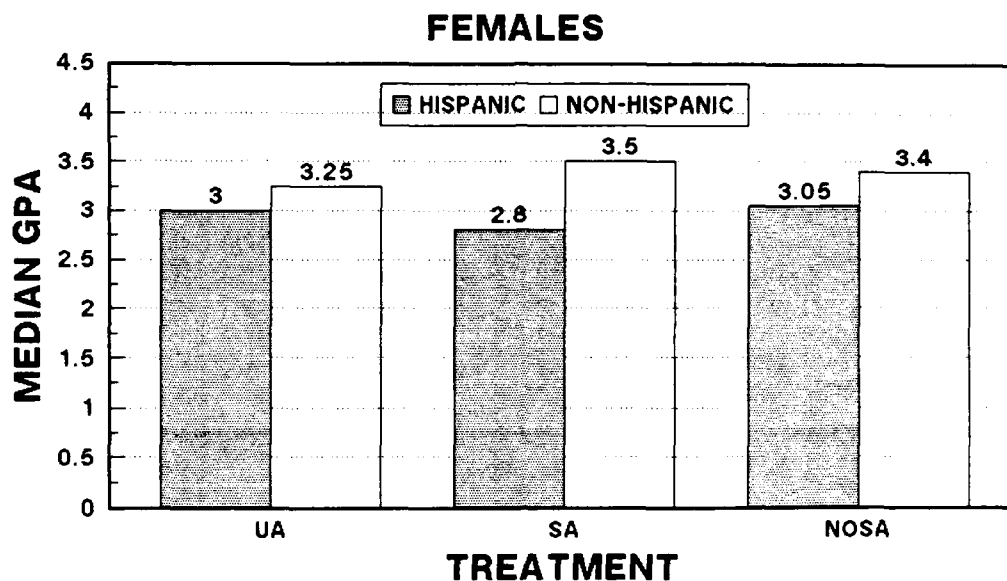
Appendix Figure M16. Standard deviations for mean risk score for each treatment for Non-Hispanic and Hispanic males.



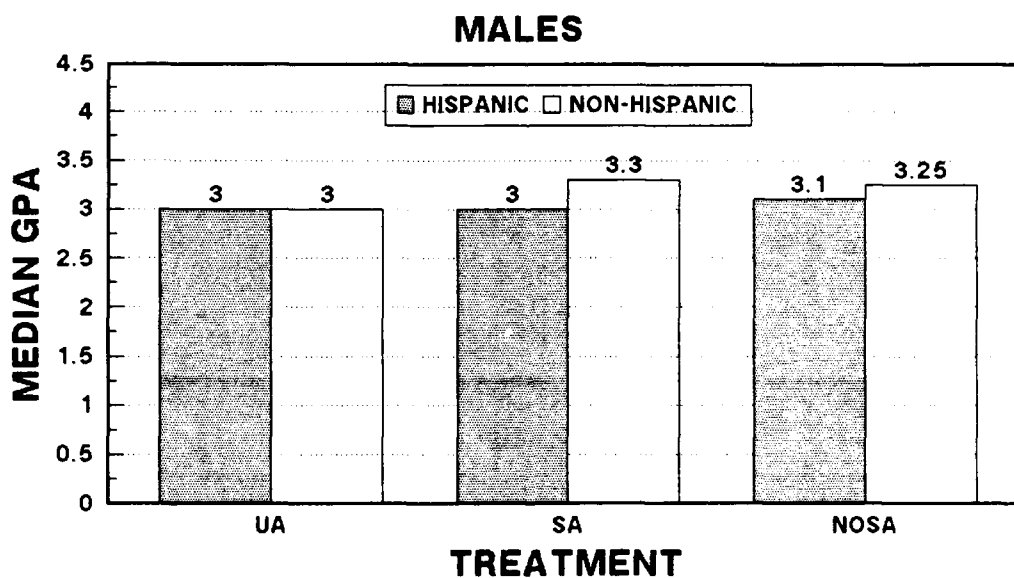
Appendix Figure M17. Mean GPA for each treatment for Non-Hispanic and Hispanic females.



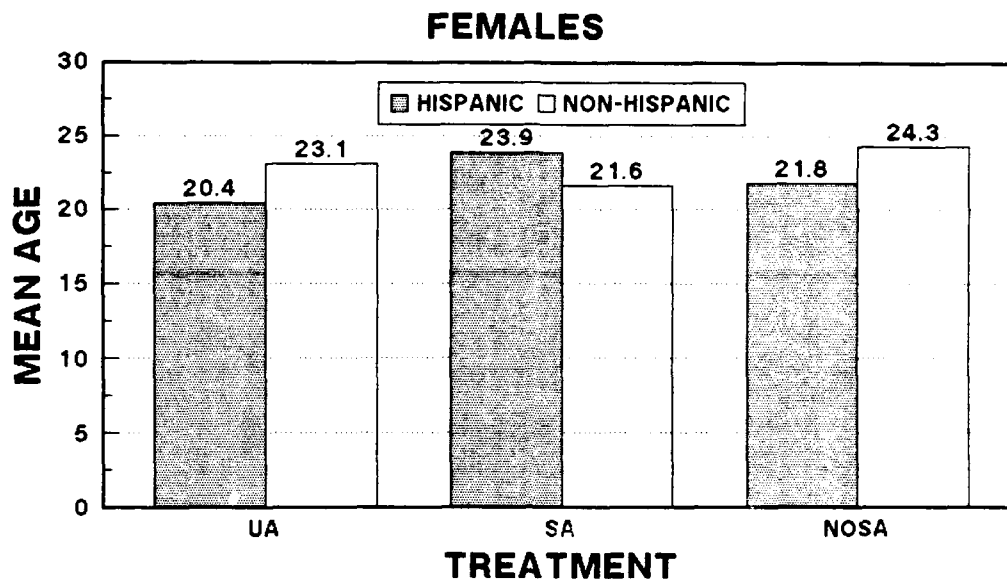
Appendix Figure M18. Mean GPA for each treatment for Non-Hispanic and Hispanic males.



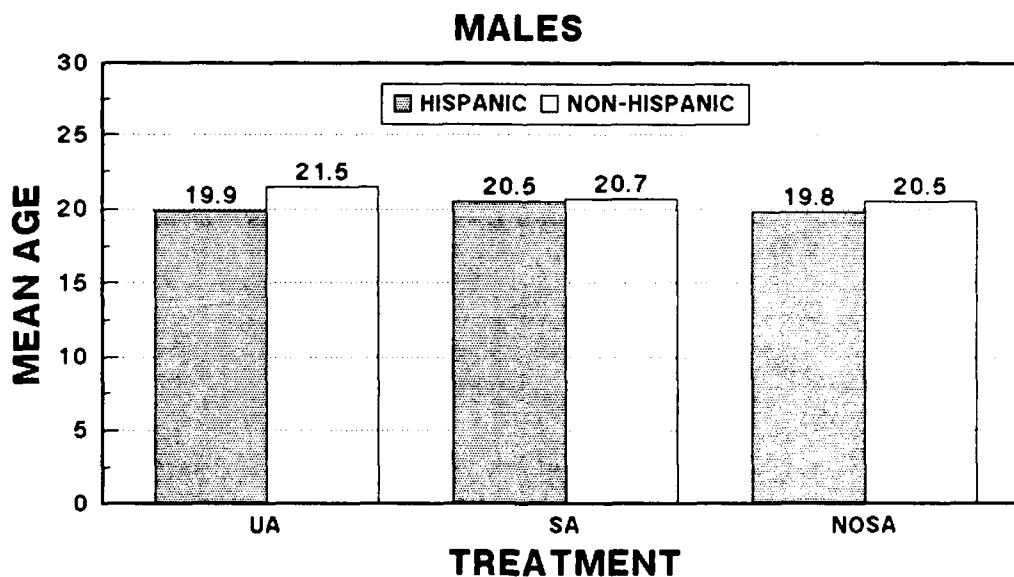
Appendix Figure M19. Median GPA for each treatment for Non-Hispanic and Hispanic females.



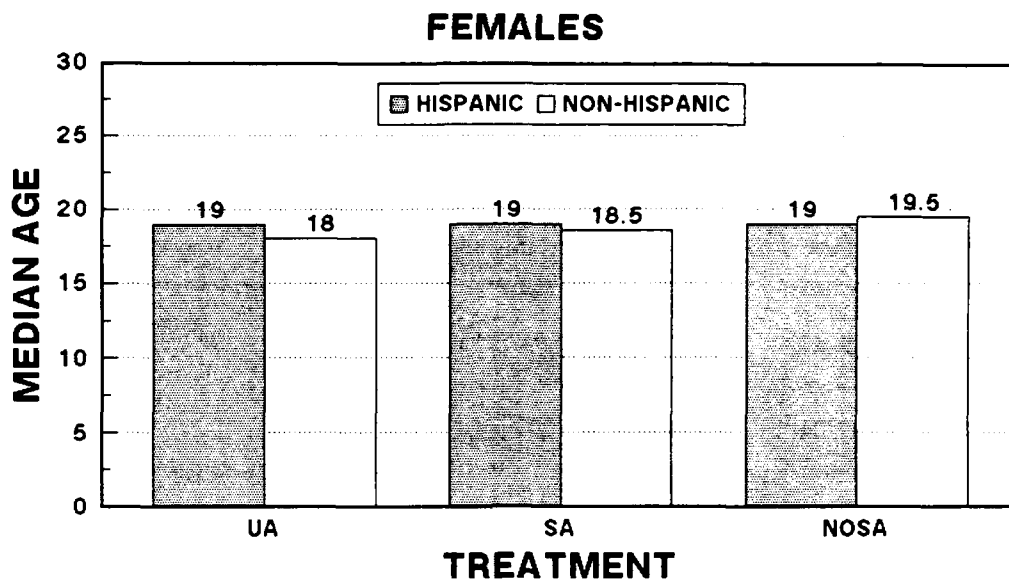
Appendix Figure M20. Median GPA for each treatment for Non-Hispanic and Hispanic males.



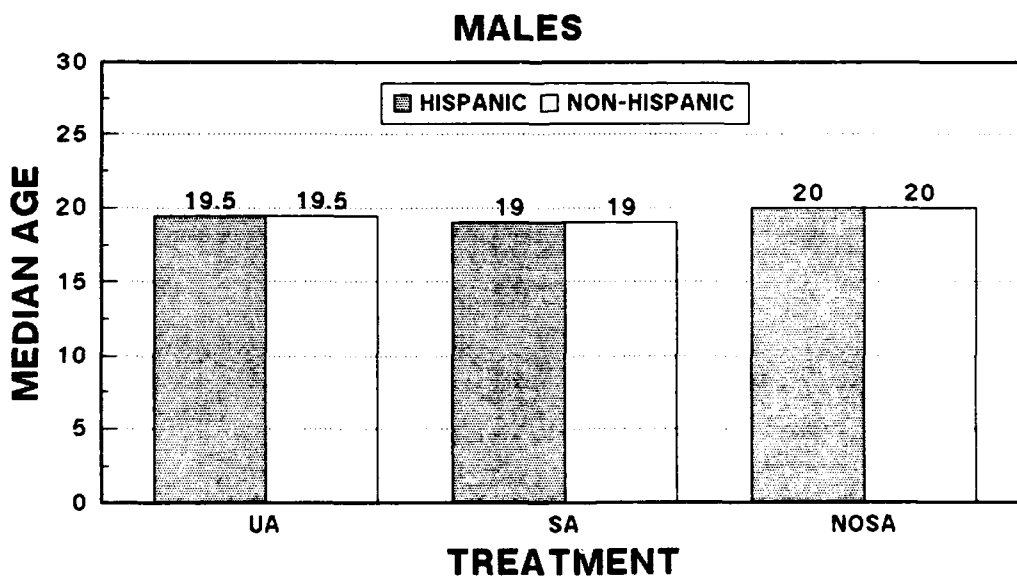
Appendix Figure M2 1. Mean age for each treatment for Non-Hispanic and Hispanic females.



Appendix Figure M22. Mean age for each treatment for Non-Hispanic and Hispanic males.



Appendix Figure M23. Median age for each treatment for Non-Hispanic and Hispanic females.



Appendix Figure M24. Median age for each treatment for Non-Hispanic and Hispanic males.